# Comparative study of harmonic and Rayleigh-Ritz procedures with applications to deflated conjugate gradients

Nicolas Venkovic, Paul Mycek, Luc Giraud, Olivier Le Maitre

CERFACS

Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique

# Comparative study of harmonic and Rayleigh-Ritz procedures with application to deflated conjugate gradients

Venkovic et al.

Technical Report TR/PA/20-3

*Publications of the Parallel Algorithms Team*

http://www.cerfacs.fr/publication/

# COMPARATIVE STUDY OF HARMONIC AND RAYLEIGH-RITZ PROCEDURES WITH APPLICATION TO DEFLATED CONJUGATE GRADIENTS

NICOLAS VENKOVIC*, PAUL MYCEK*, LUC GIRAUD†, AND OLIVIER P. LE MAÎTRE‡

**Abstract.** Harmonic Rayleigh-Ritz and Raleigh-Ritz projection techniques are compared in the context of iterative procedures to solve for small numbers of least dominant eigenvectors of large symmetric positive definite matrices. The procedures considered are (i) locally optimal conjugate gradient (CG) methods, i.e., LOBCG, (ii) thick-restart Lanczos methods, and (iii) recycled linear CG solvers, e.g., eigCG. Approaches based on principles of local optimality are adapted to enable the use of harmonic projection techniques. Upon investigating the search spaces generated by these methods, it is found that LOBCG and thick-restart Lanczos methods can be adapted, which is not the case of eigCG. Explanations are also given as to why eigCG works so well in comparison to other recycling strategies. Numerical experiments show that, while approaches based on harmonic projections consistently result in a faster convergence of eigen-residuals, they generally do not yield better convergence of the forward error of eigenvectors, until the Rayleigh quotients have converged. Then, the effect of recycling strategies is investigated on deflation for the resolution of sequences of linear systems. While non-locally optimal recycling strategies need to solve more linear systems in order to fully develop their effect on convergence, they eventually reach similar behaviors to those of locally optimal recycling procedures. While implementations based on Init-CG are robust for systems with multiple right-hand sides, this is not the case for multiple operators.

**1. Introduction.** When solving a large linear system $\mathbf{Ax} = \mathbf{b}$, approximate eigenvectors of $\mathbf{A}$ can be leveraged so as to achieve a more efficient iterative resolution [6]. Often times, this is equivalent to introducing a deflation subpsace, to which, the iterated residual is forced to remain orthogonal throughout the resolution [7, 2]. In particular, in Deflated-CG [27], the orthogonalization of the residual at each solver iteration accounts for most of the computational overcost when compared to a resolution by standard CG. Meanwhile, the smaller the gap between the actual eigenspace and the subspace spanned by the approximated vectors, the more significant of a speed-up can be achieved by deflation [8]. Floating-point arithmetic aside, if by any chance, the spectral information available is exact, then, the iterated residual does not need to be orthogonalized with respect to the deflation subspace, as it is then invariant under the action of $\mathbf{A}$. Assuming the proper initial iterate is selected, skipping the orthogonalization step leads to Init-CG [6, 21, 25, 39], which is sometimes used when the eigenvectors are sufficiently well approximated, but often does not lead to the same iteration gain as a resolution by Deflated-CG [6], depending on the stopping criterion. Irrespective of whether Deflated-CG or Init-CG is used, there is clear value in being able to extract the most accurate eigenvector approximations possible for some given computational resources.

In this work, we are particularly interested in applications where the linear system is symmetric positive definite (SPD), and in which only the least dominant (LD) eigenvectors are targeted for deflation. As most iterative eigensolvers can be recast into some sort of projection technique, see [26], Rayleigh-Ritz approximation procedures—developed after the work of Rayleigh [11] and Ritz [24]—are certainly the most common approach, see [22, 37, 26]. However, they suffer from only being optimal when applied to approximate most dominant eigenvectors, see Chap. 11.1 in [21], as well as leading to spurious approximations [29]. Harmonic Rayleigh-Ritz projection techniques were introduced and presented as better approaches to approximate interior eigen-pairs of Hermitian operators in general [13, 19, 15], i.e. not necessarily SPD. While harmonic procedures are some-

*Cerfacs, Toulouse, France (venkovic@cerfacs.fr, mycek@cerfacs.fr).
†HiePACS, Inria Bordeaux, Talence, France (luc.giraud@inria.fr).
‡CNRS, CMAP, École Polytechnique, Palaiseau, France (olivier.le-maitre@polytechnique.edu).

times preferred over classic Rayleigh-Ritz projections for LD eigenvector approximations [27, 31], some eigen-solving strategies, on the other hand, were first, or even exclusively developed using Rayleigh-Ritz projections. For example, implicitly restarted Arnoldi (IRA) methods were first developed using Rayleigh-Ritz approximations [32], before being revisited with harmonic projections [14]. Similarly, the thick-restart Lanczos method [40] was adapted to harmonic projections for the development of explicitly restarted Arnoldi algorithms [16]. Meanwhile, approaches based on optimization [3, 9, 4, 28, 10, 1] were only developed in relation to Rayleigh-Ritz projections. Finally, recycling strategies, which build search spaces directly from matrix-vector products generated by a linear iterative solver, have been considered differently depending on their characteristics. The only such method based on CG for SPD systems that recycles the search space several times throughout the resolution of a single system, i.e. eigCG [35], was based on Raleigh-Ritz projections. On the other hand, methods that generate a single search space per linear system solved, and use it to approximate the eigenvectors of the next system in a sequence—might the operator remain the same [27, 35], or not [20]—have been based on both types of projections. The question remains as to whether or not some of these strategies exclusively developed with Rayleigh-Ritz projections can be adapted to account for harmonic approximations, and if so, whether they yield better quality approximations of invariant subspaces than their counterparts based on Rayleigh-Ritz projections.

Here, we review some of these iterative eigen-solving strategies for SPD matrices, explain why they work, or not, adapt some of them to allow for harmonic approximations, and consistently compare the forward errors of the approximated eigenvectors as a function of the projection technique used. Some properties of the search spaces generated are investigated in an attempt to explain converging behaviors and lack thereof. Finally, a particular emphasis is given to recycling strategies, and their use with deflation, to solve sequences of linear systems with multiple right-hand sides, and operators. The work is organized as follows. The projection techniques used throughout this work are presented in Section 2. Then, some approaches based on optimization are presented in Section 3, where LOBCG [9, 10] is adapted to allow for harmonic approximations, albeit not for matrix pencils, and without preconditioner. Un-restarted Lanczos procedures are briefly reminded in Section 4, as forward errors of approximated eigenvectors are compared on the basis of the projection they rely on. In Section 5, different explicit (i.e., thick-) restarted strategies are reviewed while new ones are attempted, compared and analyzed. Finally, recycling strategies are investigated in Section 6, and applied with deflation, or not, to solve sequences of linear systems with different right-hand sides and operators.

**2. Projection methods.** We are interested in solving for pairs of eigenvalue $\theta$ and eigenvector $\mathbf{u}$ of symmetric positive definite matrices $\mathbf{A}$ such that $\mathbf{A}\mathbf{u} = \theta\mathbf{u}$. We assume the $n$ positive eigenvalues of $\mathbf{A}$ to be sorted in ascending order, i.e. $0 < \theta_1 \leq \cdots \leq \theta_n$, with corresponding unit eigenvectors $\mathbf{u}_i \in \mathbb{R}^n$. The first few solution pairs $\{(\theta_i, \mathbf{u}_i)\}_{i=1,2,\ldots}$ and $\{(\theta_i, \mathbf{u}_i)\}_{i=n,n-1,\ldots}$ are referred to as least and most dominant eigenpairs, respectively. Here, we consider both harmonic Rayleigh-Ritz and Raleigh-Ritz projection techniques, as they are commonly used in matrix-free iterative methods to solve for small numbers of LD or MD eigenpairs of $\mathbf{A}$.

**2.1. Rayleigh-Ritz.** Rayleigh-Ritz pairs $(\lambda, \mathbf{w})$ are approximations of eigenpairs $(\theta, \mathbf{u})$ of $\mathbf{A}$ obtained by orthogonal projection onto a subspace $\mathcal{R}(\mathbf{V})$. In particular, we search for $\lambda \in \mathbb{R}$ and $\mathbf{w} \in \mathcal{R}(\mathbf{V})$ such that the Petrov-Galerkin condition,

$$\mathbf{A}\mathbf{w} - \lambda\mathbf{w} \perp \mathcal{R}(\mathbf{V}), \tag{2.1}$$

is satisfied. Let $\mathbf{V} \in \mathbb{R}^{n\times k\leq n}$ be full column rank, so that $k$ Ritz pairs, denoted by $\{(\lambda_j, \mathbf{w}_j)\}_{j=1}^k$ with $0 < \lambda_1 \leq \cdots \leq \lambda_k$, are obtained when solving the reduced eigenvalue problem:

$$\mathbf{V}^T\mathbf{A}\mathbf{V}\hat{\mathbf{w}} = \lambda\mathbf{V}^T\mathbf{V}\hat{\mathbf{w}}, \tag{2.2}$$

$$\mathbf{V}\hat{\mathbf{w}} = \mathbf{w}. \tag{2.3}$$

A Rayleigh-Ritz procedure consists of computing and selecting only some of these solutions $(\lambda_j, \mathbf{w}_j)$. Such procedures are often considered unfit to compute approximations of the LD eigenpairs of $\mathbf{A}$. Although this difficulty can be bypassed by applying the procedure to $\mathbf{A}^{-1}$—hence mapping the eigenpairs of $\mathbf{A}$ to a sequence with reversed order—this causes other problems, namely the need to compute a factorization of the operator, or to simply solve linear systems, both of which are impractical for sufficiently large $n$.

**2.2. Harmonic Rayleigh-Ritz.** The harmonic Rayleigh-Ritz method was introduced in [13] as a mean to produce better approximations of eigenpairs with eigenvalues close to some shift $\sigma$, without having to factor the shifted operator $\mathbf{A} - \sigma\mathbf{I}$. Here, we are only interested in the LD eigenpairs of the SPD matrix $\mathbf{A}$ with eigenvalues near zero, so we simply let the shift $\sigma$ be zero. Now, one way to produce approximations of the MD eigenpairs of $\mathbf{A}^{-1}$ is to perform a Rayleigh-Ritz procedure in the subspace $\mathbf{A}\mathcal{R}(\mathbf{V})$. In particular, let $(\lambda^{-1}, \mathbf{A}\mathbf{w})$ be a Rayleigh-Ritz pair of $\mathbf{A}^{-1}$ of $\mathbf{A}\mathcal{R}(\mathbf{V})$. Every such pair satisfies the following Petrov-Galerkin condition,

$$\mathbf{A}^{-1}\mathbf{A}\mathbf{w} - \lambda^{-1}\mathbf{A}\mathbf{w} \perp \mathbf{A}\mathcal{R}(\mathbf{V}), \tag{2.4}$$

so that $k$ approximations can be obtained by solving the reduced eigenvalue problem:

$$\mathbf{V}^T\mathbf{A}^2\mathbf{V}\hat{\mathbf{w}} = \lambda\hat{\mathbf{w}}\mathbf{V}^T\mathbf{A}\mathbf{V}, \tag{2.5}$$

$$\mathbf{A}\mathbf{V}\hat{\mathbf{w}} = \mathbf{A}\mathbf{w}. \tag{2.6}$$

For every such approximation $(\lambda^{-1}, \mathbf{A}\mathbf{w})$ of eigenpair of $\mathbf{A}^{-1}$, $(\lambda, \mathbf{A}\mathbf{w})$ can be used as an approximate eigenpair of $\mathbf{A}$. However, it was argued in [13] that $(\lambda, \mathbf{w})$, the harmonic Rayleigh-Ritz pair, constitutes a better approximation. Note that, equivalently, the harmonic Rayleigh-Ritz pairs $(\lambda, \mathbf{w})$ of $\mathbf{A}$ are obtained by the projection:

$$\mathbf{w} \in \mathcal{R}(\mathbf{V}), \tag{2.7}$$

$$\mathbf{A}\mathbf{w} - \lambda\mathbf{w} \perp \mathbf{A}\mathcal{R}(\mathbf{V}). \tag{2.8}$$

It is particularly advantageous that the inverse of $\mathbf{A}$ need not be applied in order to compute $(\lambda, \mathbf{w})$. Moreover, although $(\lambda, \mathbf{w})$ is of interest, $(\rho(\mathbf{w}; \mathbf{A}), \mathbf{w})$ often provides better approximations [13, 15], where

$$\rho(\mathbf{w}; \mathbf{A}) := \frac{\mathbf{w}^T\mathbf{A}\mathbf{w}}{\mathbf{w}^T\mathbf{w}} \tag{2.9}$$

is the Rayleigh quotient. Then, harmonic Rayleigh-Ritz procedures generally consist of computing and selecting $\{(\rho(\mathbf{w}_j; \mathbf{A}), \mathbf{w}_j)\}_{j=1,2\ldots}$ to approximate some LD eigenpairs $\{(\theta_i, \mathbf{u}_i)\}_{i=1,2\ldots}$ of $\mathbf{A}$.

## 3. Approaches based on optimization.

**3.1. Rayleigh-Ritz.** Consider a sequence of iterates $\mathbf{w}^{(1)}, \mathbf{w}^{(2)}, \dots$ approximating the LD eigenvector $\mathbf{u}_1$ of $\mathbf{A}$, and let us assume $\mathbf{w}^{(r+1)} := \mathbf{w}^{(r)} + \delta^{(r)}\mathbf{p}^{(r)}$ for some search direction $\mathbf{p}^{(r)}$. Most approaches based on optimization rely on such an iteration, as well as on the fact that $\mathbf{u}_1$ minimizes the Rayleigh quotient of $\mathbf{A}$ [4, 9]. In particular, a steepest descent procedure is obtained when $\mathbf{p}^{(r)} := -\nabla\rho(\mathbf{w}^{(r)}; \mathbf{A})$ and $\delta^{(r)} := \operatorname{argmin}_\delta \rho(\mathbf{w}^{(r)} + \delta\mathbf{p}^{(r)}; \mathbf{A})$. Now, if we rather assume $\mathbf{p}^{(r)} := -\nabla\rho(\mathbf{w}^{(r)}; \mathbf{A}) + \beta^{(r)}\mathbf{p}^{(r-1)}$, the sequence follows after a nonlinear CG—indeed, the Rayleigh-quotient is not quadratic—so that different choices of $\beta^{(r)}$ yield the sought condition that $\mathbf{p}^{(r-1)}$ and $\mathbf{p}^{(r)}$ are conjugate directions, i.e. ${\mathbf{p}^{(r-1)}}^T\mathbf{A}\mathbf{p}^{(r)} = 0$. Here, we are not interested in these variants [5], but rather in those that follow from the observation that letting $\mathbf{w}^{(r+1)}_{\text{LOCG}} := \mathbf{w}^{(r)} - \delta^{(r)}_{\text{LOCG}}\nabla\rho(\mathbf{w}^{(r)}; \mathbf{A}) + \gamma^{(r)}_{\text{LOCG}}\mathbf{p}^{(r-1)}$ where $(\delta^{(r)}_{\text{LOCG}}, \gamma^{(r)}_{\text{LOCG}}) := \operatorname{argmin}_{(\delta,\gamma)} \rho(\mathbf{w}^{(r)} - \delta\nabla\rho(\mathbf{w}^{(r)}; \mathbf{A}) + \gamma\mathbf{p}^{(r-1)}; \mathbf{A})$, is such that $\rho(\mathbf{w}^{(r+1)}_{\text{LOCG}}) \leq \rho(\mathbf{w}^{(r+1)}_{\text{CG}})$ [9, 10]. Clearly, the resulting iterate is equivalent to the LD Rayleigh-Ritz vector of $\mathbf{A}$ in $\mathcal{R}(\mathbf{V}^{(r)})$ with $\mathbf{V}^{(r)} := [\mathbf{w}^{(r)}, \nabla\rho(\mathbf{w}^{(r)}; \mathbf{A}), \mathbf{p}^{(r-1)}]$. Moreover, the search space $\mathcal{R}(\mathbf{V}^{(r)})$ is equivalently spanned by $\{\mathbf{w}^{(r-1)}, \mathbf{w}^{(r)}, \tilde{\mathbf{r}}^{(r)}\}$, in which $\tilde{\mathbf{r}}^{(r)}$ denotes the eigen-residual $\tilde{\mathbf{r}}(\lambda^{(r)}, \mathbf{V}^{(r-1)}\mathbf{y}) := \mathbf{A}\mathbf{V}^{(r-1)}\mathbf{y} - \lambda^{(r)}\mathbf{V}^{(r-1)}\mathbf{y}$ of the LD eigenpair $(\lambda^{(r)}, \mathbf{y})$ of ${\mathbf{V}^{(r-1)}}^T\mathbf{A}\mathbf{V}^{(r-1)}\mathbf{y} = \lambda{\mathbf{V}^{(r-1)}}^T\mathbf{V}^{(r-1)}\mathbf{y}$. The iterative method resulting from such a sequence of Rayleigh-Ritz projections was proposed in [3, 9] for matrix pencils, and is often referred to as a locally optimal CG (LOCG)—due to the fact that it minimizes the Rayleigh quotient at least as well as any nonlinear CG scheme. Note that, while LOCG usually converges in more iterations than an unrestarted Lanczos procedure, it still can do so in less time, as it relies on search spaces of dimension no larger than three.

In [10], this method was extended and deemed the name LOBPCG so as to allow for the approximation of several eigenvectors at the same time, or improve the quality of approximation of a single eigenvector. Without preconditioning, the resulting method consists of computing Rayleigh-Ritz approximations $\mathbf{w}_1^{(r+1)}, \dots, \mathbf{w}_k^{(r+1)}$ of the LD eigenvectors $\mathbf{u}_1, \dots, \mathbf{u}_k$ in a search space

$$\text{Span}\{\mathbf{w}_1^{(r-1)}, \mathbf{w}_1^{(r)}, \dots, \mathbf{w}_k^{(r-1)}, \mathbf{w}_k^{(r)}, \tilde{\mathbf{r}}_1^{(r)}, \dots, \tilde{\mathbf{r}}_k^{(r)}\},$$

where $\tilde{\mathbf{r}}_j^{(r)}$ is the eigen-residual of the Rayleigh-Ritz pair $(\lambda_j^{(r)}, \mathbf{w}_j^{(r)})$. Since then, it has become common to refer to methods that generates search spaces using iterates $\mathbf{w}_\bullet^{(r-1)}$ and $\mathbf{w}_\bullet^{(r)}$ from subsequently generated search spaces as locally optimal methods, see [33, 34, 35] for examples.

**3.2. Harmonic Rayleigh-Ritz.** Now, just like the harmonic Rayleigh-Ritz projection is introduced as a way to bypass the difficulty related to the Rayleigh-Ritz procedure with a shifted-and-inverted operator, one can introduce a harmonic LOCG sequence as follows. First, let $\tilde{\mathbf{w}}^{(1)}, \tilde{\mathbf{w}}^{(2)}, \dots, \tilde{\mathbf{w}}^{(r)}$ be a LOCG sequence of approximations of the MD eigenvector $\mathbf{u}_1$ of $\mathbf{A}^{-1}$. This is equivalent to say that $\tilde{\mathbf{w}}^{(r+1)}$ maximizes the Rayleigh quotient of $\mathbf{A}^{-1}$ over a subspace $\mathcal{R}(\tilde{\mathbf{V}}^{(r)})$ with $\tilde{\mathbf{V}}^{(r)} := [\tilde{\mathbf{w}}^{(r-1)}, \tilde{\mathbf{w}}^{(r)}, \mathbf{A}^{-1}\tilde{\mathbf{w}}^{(r)} - (1/\lambda^{(r)})\tilde{\mathbf{w}}^{(r)}]$, where $\tilde{\mathbf{w}}^{(r)} := \tilde{\mathbf{V}}^{(r-1)}\mathbf{y}$ is obtained from the LD eigenpair $(\lambda^{(r)}, \mathbf{y})$ of $\tilde{\mathbf{V}}^{(r-1)T}\mathbf{A}^2\tilde{\mathbf{V}}^{(r-1)}\mathbf{y} = \lambda\tilde{\mathbf{V}}^{(r-1)T}\mathbf{A}\tilde{\mathbf{V}}^{(r-1)}\mathbf{y}$. Thus, by definition, $\mathbf{w}^{(r)} := \mathbf{A}^{-1}\tilde{\mathbf{w}}^{(r)}$ is the LD harmonic Rayleigh-Ritz vector of $\mathbf{A}$ in $\mathcal{R}(\mathbf{V}^{(r)})$, where $\mathbf{V}^{(r)} := \mathbf{A}^{-1}\tilde{\mathbf{V}}^{(r)} = [\mathbf{w}^{(r-1)}, \mathbf{w}^{(r)}, \mathbf{A}^{-1}\mathbf{w}^{(r)} - (1/\lambda^{(r)})\mathbf{w}^{(r)}]$. Our numerical results show that a procedure based on such a sequence of harmonic projections converges very fast. However, the presence of $\mathbf{A}^{-1}$ in the last vector spanning the search space defeats the purpose of harmonic projections as a mean to bypass the need to shift-and-invert the operator. To circumvent this difficulty, we note that

$$\mathbf{A}^{-1}\mathbf{w}^{(r)} - (1/\lambda^{(r)})\mathbf{w}^{(r)} = -(1/\lambda^{(r)})\mathbf{A}^{-1}\tilde{\mathbf{r}}^{(r)},$$

where the harmonic eigen-residual $\tilde{\mathbf{r}}^{(r)} := \mathbf{A}\mathbf{w}^{(r)} - \lambda^{(r)}\mathbf{w}^{(r)}$ is computed with the harmonic Rayleigh-Ritz value $\lambda^{(r)}$—not the Rayleigh quotient. Note that this does not resolve our problem, as we still need to compute $\mathbf{A}^{-1}\tilde{\mathbf{r}}^{(r)}$. To bypass this difficulty, we rather consider $\mathrm{Span}\{\mathbf{w}^{(r-1)}, \mathbf{w}^{(r)}, \tilde{\mathbf{z}}^{(r)}\}$ as a search space for the harmonic projection, where $\tilde{\mathbf{z}}^{(r)} := \mathbf{V}^{(r-1)}({\mathbf{V}^{(r-1)}}^T\mathbf{A}\mathbf{V}^{(r-1)})^{-1}{\mathbf{V}^{(r-1)}}^T\tilde{\mathbf{r}}^{(r)}$ approximates $\mathbf{A}^{-1}\tilde{\mathbf{r}}^{(r)}$ in the previously generated search space. The resulting procedure is straightforwardly generalized to enable the computation of several approximate eigenvectors at the same time.

**3.3. Numerical results.** As shown in Fig. 1, the harmonic LOBCG procedure yields faster convergence of eigen-residuals. However, the forward errors of the approximated eigenvectors only start converging faster than those of Rayleigh-Ritz LOBCG once the Rayleigh quotients have converged.

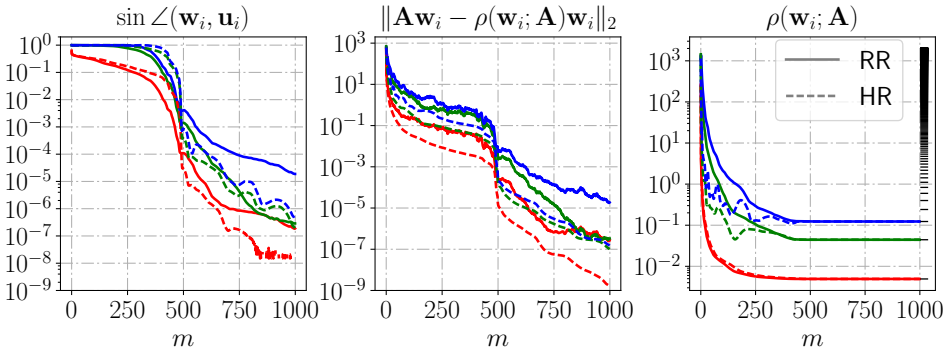


Figure 1: Results of LOBCG procedures with RR and HR approximations. The LD eigenvectors $\mathbf{u}_i$ of $\mathbf{A}$ are approximated by $\mathbf{w}_i$ for $i = 1$ (—), $i = 2$ (—), $i = 3$ (—). The black ticks on the third column represent the spectrum $\mathrm{Sp}(\mathbf{A})$.

**4. Unrestarted Lanczos procedures.** The unrestarted Lanczos procedure constructs an orthonormal basis $\mathbf{v}_1, \ldots, \mathbf{v}_m$ of the Krylov subspace of $\mathbf{A}$ generated by some unit vector $\mathbf{v}_1$:

$$\mathrm{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_m\} = \mathcal{K}_m(\mathbf{A}, \mathbf{v}_1) := \mathrm{Span}\{\mathbf{v}_1, \mathbf{A}\mathbf{v}_1, \ldots, \mathbf{A}^{m-1}\mathbf{v}_1\}.$$

This basis is stored in the columns of $\mathbf{V}_m$, which admits the following relation:

$$\mathbf{A}\mathbf{V}_m = \mathbf{V}_m\mathbf{V}_m^T\mathbf{A}\mathbf{V}_m + \beta_m\mathbf{v}_{m+1}\mathbf{e}_m^T,$$

where $\mathbf{e}_m$ is the last column of the $m$-dimensional identity matrix, and $\beta_m := \mathbf{v}_m^T\mathbf{A}\mathbf{v}_{m+1}$.

Note that for all Rayleigh-Ritz pairs $(\lambda, \mathbf{w})$ of $\mathbf{A}$ in $\mathcal{R}(\mathbf{V}_m)$, we have:

$$\tilde{\mathbf{r}}(\lambda, \mathbf{w}) := \mathbf{A}\mathbf{w} - \lambda\mathbf{w} = \mathbf{A}\mathbf{V}_m\mathbf{y} - \lambda\mathbf{V}_m\mathbf{y} = \beta_m(\mathbf{e}_m^T\mathbf{y})\mathbf{v}_{m+1},$$

where $\mathbf{w} := \mathbf{V}_m\mathbf{y}$ and $(\lambda, \mathbf{y})$ is an eigenpair of $\mathbf{V}_m^T\mathbf{A}\mathbf{V}_m$.

**4.1. Numerical results.** See results in Fig. 2. Although the harmonic procedure leads to faster convergence of eigen-residuals and Rayleigh quotients, it does not result in a faster convergence of the eigenvectors and the subspace they span.

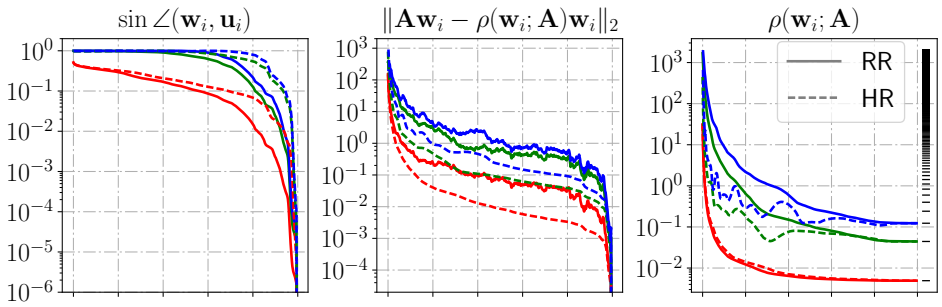


Figure 2: Results of un-restarted Lanczos procedures with RR and HR approximations. The LD eigenvectors $\mathbf{u}_i$ of $\mathbf{A}$ are approximated by $\mathbf{w}_i$ for $i = 1$ (—), $i = 2$ (—), $i = 3$ (—). The black ticks on the third column represent the spectrum $\mathrm{Sp}(\mathbf{A})$.

**5. Restarted Lanczos procedures.** For large matrices, high-dimensional search spaces are often needed in order for the Rayleigh-Ritz vectors to precisely approximate several eigenvectors of $\mathbf{A}$. When the search space is generated by a Lanczos procedure, this requires storing a large basis so as to compute the Rayleigh-Ritz vectors using Eq. (2.3), after solving Eq. (2.2). Until then, the Lanczos vectors can be stored on disk. However, due to the effect of floating-point arithmetic, the computed vectors tend to lose their orthogonality as the dimension of the spanned Krylov subspace increases. This phenomenon, explained in [17, 18], requires some re-orthogonalization to prevent collateral effects on the Rayleigh-Ritz vectors. Different strategies exist to re-orthogonalize: full re-orthogonalization, which is the most computationally demanding; to which, partial [30], and selective [23] schemes were proposed as alternatives. Irrespective of the strategy selected, the cost of re-orthogonalization remains an issue for high-dimensional problems, often requiring Lanczos vectors to be stored on-core for a faster execution, hence limiting the range of suitable applications for un-restarted algorithms. Restarting strategies allow for a better use of resources than un-restarted algorithms [26].

**5.1. Thick-restarts.** Here, we exclusively consider strategies in which the search space is explicitly restarted with several eigenvector approximations. While this approach was deemed the term *thick-restart* by Wu and Simon [40], it was not always the case when similar ideas were developed [12, 38]. We define thick-restart methods as follows. First, the Lanczos vectors of a Krylov subspace $\mathcal{K}_\bullet(\mathbf{A}, \mathbf{v}_1)$ are progressively generated to form search spaces of increasing dimension, each of which can be used to calculate eigenvector approximations. Once the dimension of the search space reaches a certain size $k + \ell$, its $k$ LD eigenvector approximations $\mathbf{w}_1, \ldots, \mathbf{w}_k$ are used as the basis of a new, *restarted* subspace. From here on, a new vector $\hat{\mathbf{v}}_1 \perp \mathrm{Span}\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ is used to generate new Lanczos vectors $\hat{\mathbf{v}}_2, \ldots, \hat{\mathbf{v}}_j$, by letting each $\hat{\mathbf{v}}_j$ be $\mathbf{A}\hat{\mathbf{v}}_{j-1}$ orthogonalized against $\mathbf{w}_1, \ldots, \mathbf{w}_k, \hat{\mathbf{v}}_1, \ldots, \hat{\mathbf{v}}_{j-1}$. Once $\ell$ new Lanczos vectors have been generated, $k$ new eigenvector approximations are computed in the current search space $\mathcal{S}_{\text{TR-Lan}}(\{\mathbf{w}_i\}_{i=1}^k, \hat{\mathbf{v}}_1, \ell)$ defined by

$$\mathcal{S}_{\text{TR-Lan}}(\{\mathbf{w}_i\}_{i=1}^k, \hat{\mathbf{v}}_1, j) := \mathrm{Span}\{\mathbf{w}_1, \ldots, \mathbf{w}_k, \hat{\mathbf{v}}_1, \mathbf{A}\hat{\mathbf{v}}_1, \ldots, \mathbf{A}^{j-1}\hat{\mathbf{v}}_1\} \tag{5.1}$$

for $1 \leq j \leq \ell$. The search space is restarted again with these approximations and some $\hat{\mathbf{v}}_1$, yet to be defined. This process is repeated over and over again, and the overall number of Lanczos vectors

generated is denoted by $m$. For all $m > k + \ell$, the current search space can be put in the form of Eq. (5.1). Note that, as the orthogonalization of $\hat{\mathbf{v}}_2, \ldots, \hat{\mathbf{v}}_j$ is done by a Gram-Schmidt procedure, it can simplify to short recurrence relations depending on how $\hat{\mathbf{v}}_1$ is defined for the given projection technique. Moreover, the choice of $\hat{\mathbf{v}}_1$ also strongly influences some properties of the current search space which are responsible for the convergence of the restarted procedure.

**5.1.1. Rayleigh-Ritz.** Consider the case addressed by Wu and Simon [40], in which the approximations $(\lambda_1, \mathbf{w}_1), \ldots, (\lambda_k, \mathbf{w}_k)$ are Rayleigh-Ritz pairs of $\mathbf{A}$ in the current search space. Just before the search space is restarted for the first time, by property of Rayleigh-Ritz pairs in Krylov subspaces, we have

$$\mathbf{A}\mathbf{w}_i = \lambda_i \mathbf{w}_i + \beta_{k+\ell}(\mathbf{e}_{k+\ell}^T \mathbf{y}_i)\mathbf{v}_{k+\ell+1} \tag{5.2}$$

where $\mathbf{w}_i := [\mathbf{v}_1, \ldots, \mathbf{v}_{k+\ell}]\mathbf{y}_i$ for $1 \leq i \leq k$. Then, letting $\hat{\mathbf{v}}_1 := \mathbf{v}_{k+\ell+1}$ has important consequences on the procedure. First, the computation of $\hat{\mathbf{v}}_2, \ldots, \hat{\mathbf{v}}_j$ is simplified to the evaluation of short recurrence relations, see [40]. We then have:

$$\mathbf{A}\hat{\mathbf{v}}_1 = \hat{\alpha}_1 \hat{\mathbf{v}}_1 + \hat{\beta}_1 \hat{\mathbf{v}}_2 + \sum_{i=1}^{k} \beta_{k+\ell}(\mathbf{e}_{k+\ell}^T \mathbf{y}_i)\mathbf{w}_i, \tag{5.3}$$

$$\mathbf{A}\hat{\mathbf{v}}_j = \hat{\alpha}_j \hat{\mathbf{v}}_j + \hat{\beta}_j \hat{\mathbf{v}}_{j+1} + \hat{\beta}_{j-1} \hat{\mathbf{v}}_{j-1} \quad \text{for} \quad j > 1, \tag{5.4}$$

where $\hat{\alpha}_j := \hat{\mathbf{v}}_j^T \mathbf{A}\hat{\mathbf{v}}_j$ and $\hat{\beta}_j := \hat{\mathbf{v}}_{j+1}^T \mathbf{A}\hat{\mathbf{v}}_j$, so that $\mathbf{A}[\mathbf{w}_1, \ldots, \mathbf{w}_k, \hat{\mathbf{v}}_1, \ldots, \hat{\mathbf{v}}_j]$ can be put in the form of a rank-1 update which, in turn, implies that the current search space $\mathcal{S}_{\text{TR-Lan}}$ is, and remains a Krylov subspace as long as $\hat{\mathbf{v}}_1 := \hat{\mathbf{v}}_{\ell+1}$ for all the next restarts. Then, once passed the first restart, all the eigenvector approximations are such that

$$\mathbf{A}\mathbf{w}_i = \lambda_i \mathbf{w}_i + \hat{\beta}_\ell(\mathbf{e}_{k+\ell}^T \mathbf{y}_i)\hat{\mathbf{v}}_{\ell+1}. \tag{5.5}$$

An important observation is that Eqs. (5.1), (5.2) and (5.5) imply

$$\mathcal{S}_{\text{TR-Lan}}(\{\mathbf{w}_r\}_{r=1}^k, \hat{\mathbf{v}}_1, j) = \text{Span}\{\mathbf{w}_1, \ldots, \mathbf{w}_k, \mathbf{A}\mathbf{w}_i, \ldots, \mathbf{A}^j \mathbf{w}_i\} \quad \text{for} \quad 1 \leq i \leq k. \tag{5.6}$$

This, in turn, implies that the current $(k + j)$-dimensional search space $\mathcal{S}_{\text{TR-Lan}}$ contains all the $(j + 1)$-dimensional Krylov subspaces of $\mathbf{A}$ generated by each of the eigenvector approximations $\mathbf{w}_1, \ldots, \mathbf{w}_k$:

$$\mathcal{K}_{j+1}(\mathbf{A}, \mathbf{w}_i) \subset \mathcal{S}_{\text{TR-Lan}}(\{\mathbf{w}_r\}_{r=1}^k, \hat{\mathbf{v}}_1, j) \quad \text{for} \quad 1 \leq i \leq k. \tag{5.7}$$

This property plays an important role in explaining the effectiveness, and lack thereof, of restarted methods [16]. Therefore, all the other methods presented in this work are investigated to tell whether or not they admit a property similar to Eq. (5.7).

**5.1.2. Harmonic Rayleigh-Ritz.** Clearly, if there exists $\tilde{\mathbf{r}}$ such that $\mathbf{A}\mathbf{w}_i = \lambda_i \mathbf{w}_i + \gamma_i \tilde{\mathbf{r}}$ with some constant $\gamma_i$ for all $1 \leq i \leq k$, then, restarting with $\hat{\mathbf{v}}_1 := \tilde{\mathbf{r}}$ ensures that the newly generated search space satisfies Eq. (5.7). For the case of Rayleigh-Ritz approximations, the existence of such a $\tilde{\mathbf{r}}$ is guaranteed as long as the search space is a Krylov subspace. Now, when $(\lambda_1, \mathbf{w}_1), \ldots, (\lambda_k, \mathbf{w}_k)$ are harmonic pairs of $\mathbf{A}$, such a $\tilde{\mathbf{r}}$ also exists, under the same condition.

*Proof.* Let $\{(\lambda_i, \mathbf{w}_i)\}_{i=1}^k$ be harmonic Rayleigh-Ritz pairs of $\mathbf{A}$ in $\mathcal{K}_m(\mathbf{A}, \mathbf{v}_1)$. The matrix $\mathbf{A}$ being positive definite, we have $\mathcal{K}_m(\mathbf{A}, \mathbf{v}_1) = \mathcal{K}_m(\mathbf{A}^{-1}, \mathbf{A}^{m-1}\mathbf{v}_1)$. By definition of harmonic pairs, $\{(\lambda_i^{-1}, \mathbf{A}\mathbf{w}_i)\}_{i=1}^k$ are Rayleigh-Ritz pairs of $\mathbf{A}^{-1}$ in $\mathcal{K}_m(\mathbf{A}^{-1}, \mathbf{A}^{m-1}\mathbf{v}_1)$. Then, by property of Rayleigh-Ritz approximations in Krylov subspaces, each eigen-residual $\mathbf{A}^{-1}\mathbf{A}\mathbf{w}_i - \lambda_i^{-1}\mathbf{A}\mathbf{w}_i$ can be expressed as a multiple of the same unit vector $\tilde{\mathbf{r}}$, say $-(\gamma_i/\lambda_i)\tilde{\mathbf{r}}$ for some $\gamma_i$. ∎

Therefore, since the search space is always Krylov prior to the first restart, letting $\hat{\mathbf{v}}_1 := \tilde{\mathbf{r}}$, as done in [16], ensures Eq. (5.7) is satisfied between the first and second restarts. Then, the orthogonalization leading to the computation of $\hat{\mathbf{v}}_2, \ldots, \hat{\mathbf{v}}_j$ simplifies to short recurrence relations, and, similarly as for Rayleigh-Ritz projections, $\mathbf{A}\hat{\mathbf{v}}_2, \ldots, \mathbf{A}\hat{\mathbf{v}}_j$ are given by Eq. (5.4), while

$$\mathbf{A}\hat{\mathbf{v}}_1 = \hat{\alpha}_1\hat{\mathbf{v}}_1 + \hat{\beta}_1\hat{\mathbf{v}}_2 + \sum_{i=1}^{k}(\lambda_i\tilde{\mathbf{r}}^T\mathbf{w}_i + \gamma_i)\mathbf{w}_i.$$

As a result, the current search space $\mathcal{S}_{\text{TR-Lan}}$ is, and remains a Krylov subspace as long as $\hat{\mathbf{v}}_1 := \tilde{\mathbf{r}}$ for all the next restarts, which in turn implies that Eq. (5.7) is satisfied throughout the procedure.

On the other hand, performing the first restart with $\hat{\mathbf{v}}_1 := \mathbf{v}_{k+\ell+1}$ has for consequence that, in general, the new search space does not contain any Krylov subspace of $\mathbf{A}$ of dimension $d > 1$ generated by any of the approximations $\mathbf{w}_1, \ldots, \mathbf{w}_k$. It also does not contain any of the vectors $\mathbf{A}^d\mathbf{w}_i$ for $d > 0$.

*Proof.* First, for $\mathcal{K}_2(\mathbf{A}, \mathbf{w}_i)$ to be contained in $\mathcal{S}_{\text{TR-Lan}}(\{\mathbf{w}_r\}_{r=1}^k, \mathbf{v}_{k+\ell+1}, j)$, there must exist some constants $a_1^{(i)}, \ldots, a_k^{(i)}$ and $b_1^{(i)}, \ldots, b_j^{(i)}$ such that

$$\mathbf{A}\mathbf{w}_i = \sum_{r=1}^{k} a_r^{(i)}\mathbf{w}_r + \sum_{s=1}^{j} b_s^{(i)}\hat{\mathbf{v}}_s \tag{5.8}$$

is satisfied. Let $\mathbf{V}_{k+\ell} := [\mathbf{v}_1, \ldots, \mathbf{v}_{k+\ell}]$, so that every harmonic eigenvector approximation can be expressed as $\mathbf{w}_i = \mathbf{V}_{k+\ell}\mathbf{y}_i$. Then, since the columns of $\mathbf{V}_{k+\ell}$ span a Krylov subspace, Eq. (5.8) can be recast in

$$\left(\mathbf{V}_{k+\ell}\mathbf{T}_{k+\ell} + \beta_{k+\ell}\mathbf{v}_{k+\ell+1}\mathbf{e}_{k+\ell}^T\right)\mathbf{y}_i = \sum_{r=1}^{k} a_r^{(i)}\mathbf{V}_{k+\ell}\mathbf{y}_r + \sum_{s=1}^{j} b_s^{(i)}\hat{\mathbf{v}}_s$$

$$\mathbf{V}_{k+\ell}\mathbf{T}_{k+\ell}\mathbf{y}_i + \beta_{k+\ell}(\mathbf{e}_{k+\ell}^T\mathbf{y}_i)\mathbf{v}_{k+\ell+1} = \mathbf{V}_{k+\ell}\sum_{r=1}^{k} a_r^{(i)}\mathbf{y}_r + \sum_{s=1}^{j} b_s^{(i)}\hat{\mathbf{v}}_s$$

where $\mathbf{T}_{k+\ell} := \mathbf{V}_{k+\ell}^T\mathbf{A}\mathbf{V}_{k+\ell}$. However, $\mathbf{v}_{k+\ell+1}$ and $\hat{\mathbf{v}}_s$ are orthogonal to $\mathcal{R}(\mathbf{V}_{k+\ell})$ by construction, so that we should have

$$\mathbf{T}_{k+\ell}\mathbf{y}_i = \sum_{r=1}^{k} a_r^{(i)}\mathbf{y}_r,$$

where $\mathbf{y}_1, \ldots, \mathbf{y}_k$ are eigenvectors of the matrix pencil $(\mathbf{V}_{k+\ell}^T\mathbf{A}^2\mathbf{V}_{k+\ell}, \mathbf{T}_{k+\ell})$. Note that, in general, these eigenvectors span a different subspace than the eigenvectors of $\mathbf{T}_{k+\ell}$. Moreover, since $\mathbf{A}$ is definite, $\ell > 1$ implies that $\mathcal{R}([\mathbf{y}_1, \ldots, \mathbf{y}_k])$ does not contain $\mathcal{R}(\mathbf{T}_{k+\ell})$. Thus, we can generally not find the constants $a_1^{(i)}, \ldots, a_k^{(i)}$ such that Eq. (5.8) is satisfied, in which case, none of the subspaces $\mathcal{K}_d(\mathbf{A}, \mathbf{w}_i)$ is included in the search space $\mathcal{S}_{\text{TR-Lan}}$ for $d > 1$.

Second, if $\mathbf{A}^d\mathbf{w}_i \notin \mathcal{R}(\hat{\mathbf{V}})$, where $\hat{\mathbf{V}} := [\mathbf{w}_1, \ldots, \mathbf{w}_k, \hat{\mathbf{v}}_1, \ldots, \hat{\mathbf{v}}_j]$, then $\mathbf{A}^{d+1}\mathbf{w}_i \notin \mathcal{R}(\hat{\mathbf{V}})$. Indeed, if there exists $\hat{\mathbf{z}}$ such that $\mathbf{A}^{d+1}\mathbf{w}_i = \hat{\mathbf{V}}\hat{\mathbf{z}}$, then, there must exist $\mathbf{z} := (\hat{\mathbf{V}}^T\mathbf{A}\hat{\mathbf{V}})^{-1}\hat{\mathbf{z}}$ such that $\mathbf{A}^d\mathbf{w}_i = \hat{\mathbf{V}}\mathbf{z}$. Therefore, $\mathbf{A}\mathbf{w}_i \notin \mathcal{S}_{\text{TR-Lan}}$ implies $\mathbf{A}^2\mathbf{w}_i \notin \mathcal{S}_{\text{TR-Lan}}$, which implies $\mathbf{A}^3\mathbf{w}_i \notin \mathcal{S}_{\text{TR-Lan}}$, and so on. ∎

Letting $\hat{\mathbf{v}}_1 := \mathbf{v}_{k+\ell+1}$ also has for consequence that the orthogonalization leading to the computation of $\hat{\mathbf{v}}_2, \ldots, \hat{\mathbf{v}}_j$ does not simplify. Then, the new search space is not Krylov, and Eq. (5.7) generally still does not hold when the next restarts are done with $\hat{\mathbf{v}}_1 := \hat{\mathbf{v}}_{\ell+1}$.

**5.2. Locally optimal thick-restarts.** A thick-restart procedure based on Rayleigh-Ritz approximations of a single eigenvector with $(k, \ell) = (1, 1)$ leads to a sequence which is equivalently given by a steepest descent approach to minimize a Rayleigh quotient over a subspace $\text{Span}\{\mathbf{w}^{(r)}, \tilde{\mathbf{r}}^{(r)}\}$, which does *not* include the previous iterate $\mathbf{w}^{(r-1)}$. Since a locally optimal iterate would be obtained from a subspace which does include this previous approximate, we now consider thick-restart strategies which incorporate eigenvector approximations from subsequent subspaces when generating a new search space. Locally optimal thick-restarts were used in [36] for a generalized Davidson method. We define locally optimal thick-restarts as follows. Similarly as for the regular thick-restarts presented in Section 5.1, we start by generating Lanczos vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots$ until the spanned Krylov subspace reaches a dimension $2k + \ell$. Then, a first restart is achieved with, still, the $k$ LD eigenvector approximations $\mathbf{w}_1, \ldots, \mathbf{w}_k$ in $\mathcal{R}(\mathbf{V}_{2k+\ell})$, where $\mathbf{V}_\bullet := [\mathbf{v}_1, \ldots, \mathbf{v}_\bullet]$, along with the $k$ LD approximations $\overline{\mathbf{w}}_1, \ldots, \overline{\mathbf{w}}_k$ in $\mathcal{R}(\mathbf{V}_{2k+\ell-1})$. Then, a restart vector $\hat{\mathbf{v}}_1 \perp \text{Span}\{\mathbf{w}_1, \ldots, \mathbf{w}_k, \overline{\mathbf{w}}_1, \ldots, \overline{\mathbf{w}}_k\}$ is defined, and a basis of the subspace

$$\begin{aligned} \mathcal{S}_{\text{LO-TR-Lan}}(\{(\mathbf{w}_i, \overline{\mathbf{w}}_i)\}_{i=1}^k, \hat{\mathbf{v}}_1, j) &:= \text{Span}\{\mathbf{w}_1, \overline{\mathbf{w}}_1, \ldots, \mathbf{w}_k, \overline{\mathbf{w}}_k, \hat{\mathbf{v}}_1, \mathbf{A}\hat{\mathbf{v}}_1, \ldots, \mathbf{A}^{j-1}\hat{\mathbf{v}}_1\} \\ &= \text{Span}\{\mathbf{w}_1, \overline{\mathbf{w}}_1, \ldots, \mathbf{w}_k, \overline{\mathbf{w}}_k, \hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2, \ldots, \hat{\mathbf{v}}_j\} \end{aligned} \tag{5.9}$$

is completed with some vectors $\hat{\mathbf{v}}_j$, each of which obtained by orthogonalization of $\mathbf{A}\hat{\mathbf{v}}_{j-1}$ against $\mathbf{w}_1, \overline{\mathbf{w}}_1, \ldots, \mathbf{w}_k, \overline{\mathbf{w}}_k, \hat{\mathbf{v}}_1, \ldots, \hat{\mathbf{v}}_{j-1}$. Once $\ell$ new Lanczos vectors have been generated, the search space is restarted again with $k$ eigenvector approximations from $\mathcal{S}_{\text{LO-TR-Lan}}(\{(\mathbf{w}_i, \overline{\mathbf{w}}_i)\}_{i=1}^k, \hat{\mathbf{v}}_1, \ell)$, $k$ others in $\mathcal{S}_{\text{LO-TR-Lan}}(\{(\mathbf{w}_i, \overline{\mathbf{w}}_i)\}_{i=1}^k, \hat{\mathbf{v}}_1, \ell - 1)$, and a restart vector $\hat{\mathbf{v}}_1$. As this process is repeated, the overall number of Lanczos vectors generated is denoted by $m$, so that the current search space can be put in the form of Eq. (5.9) for all $m > k + \ell$. Similarly as for the standard thick-restart, the choice of $\hat{\mathbf{v}}_1$ bears important consequences on the behavior of this procedure.

In practice, the pairs $(\mathbf{w}_1, \overline{\mathbf{w}}_1)$ through $(\mathbf{w}_k, \overline{\mathbf{w}}_k)$ tend to be co-linear. Although this can be interpreted as a sign of convergence of the procedure, when building approximations with unit norms, we actually observe $\mathbf{w}_i^T\overline{\mathbf{w}}_j \approx \delta_{ij}$ directly after the first restarts. So that, if convergence there is, this is a rather slow one. These near linear dependences can cause some loss of rank to appear in the matrices of the reduced eigen-problem, eventually leading to spurious approximations. A simple solution to this problem is obtained when substituting $\mathbf{w}_1, \overline{\mathbf{w}}_1, \ldots, \mathbf{w}_k, \overline{\mathbf{w}}_k$ by all the $q$ eigenvector approximations $\hat{\mathbf{w}}_1, \ldots, \hat{\mathbf{w}}_q$ computed in a subspace

$$\mathcal{R}(\mathbf{V}_{k+\ell}\mathbf{Q}) = \text{Span}\{\mathbf{w}_1, \overline{\mathbf{w}}_1, \ldots, \mathbf{w}_k, \overline{\mathbf{w}}_k\},$$

where $\mathbf{Q}$ has full column rank $q$ with $k < q \leq 2k$. More particularly, the columns of $\mathbf{Q}$ form an orthonormal basis of

$$\mathcal{R}(\mathbf{Q}) = \text{Span}\left\{\mathbf{y}_1, \begin{bmatrix}\overline{\mathbf{y}}_1 \\ 0\end{bmatrix}, \ldots, \mathbf{y}_k, \begin{bmatrix}\overline{\mathbf{y}}_k \\ 0\end{bmatrix}\right\},$$

where $\mathbf{y}_1, \dots, \mathbf{y}_k$ and $\overline{\mathbf{y}}_1, \dots, \overline{\mathbf{y}}_k$ are the reduced eigenvectors which, in the case of the first restart, are such that $\mathbf{w}_i = \mathbf{V}_{2k+\ell}\mathbf{y}_i$ and $\overline{\mathbf{w}}_i = \mathbf{V}_{2k+\ell-1}\overline{\mathbf{y}}_i$, respectively. This solution, which was adopted in [35], allows for an efficient orthogonalization of $\mathbf{w}_1, \overline{\mathbf{w}}_1, \dots, \mathbf{w}_k, \overline{\mathbf{w}}_k$, using only $(2k+\ell)$-dimensional vectors.

**5.2.1. Rayleigh-Ritz.** First, we consider the case in which $\mathbf{w}_1, \dots, \mathbf{w}_k$ and $\overline{\mathbf{w}}_1, \dots, \overline{\mathbf{w}}_k$ are Rayleigh-Ritz approximations in $\mathcal{R}(\mathbf{V}_{2k+\ell})$ and $\mathcal{R}(\mathbf{V}_{2k+\ell-1})$, respectively. An important observation is that, irrespective of $\hat{\mathbf{v}}_1$, there is at least one vector $\hat{\mathbf{w}}_i$ among $\hat{\mathbf{w}}_1, \dots, \hat{\mathbf{w}}_q$, such that

$$\mathcal{K}_d(\mathbf{A}, \hat{\mathbf{w}}_i) \not\subset \mathcal{S}_{\text{LO-TR-Lan}}(\{\hat{\mathbf{w}}_r\}_{r=1}^q, \hat{\mathbf{v}}_1, j) \text{ for all } d > 1. \tag{5.10}$$

*Proof.* Let $\hat{\mathbf{w}}_i := \mathbf{V}_{2k+\ell}\mathbf{Q}\hat{\mathbf{q}}_i$ for $1 \leq i \leq q$, where $(\lambda_i, \hat{\mathbf{q}}_i)$ is an eigen-pair of $\mathbf{Q}^T\mathbf{V}_{2k+\ell}^T\mathbf{A}\mathbf{Q}\mathbf{V}_{2k+\ell}$. For a Krylov subspace of dimension $d > 1$ generated by $\hat{\mathbf{w}}_i$, to be contained in $\mathcal{S}_{\text{LO-TR-Lan}}$, there must exist some constants $a_1^{(i)}, \dots, a_q^{(i)}, b_1^{(i)}, \dots, b_j^{(i)}$ such that

$$\mathbf{A}\hat{\mathbf{w}}_i = \sum_{r=1}^{q} a_r^{(j)}\hat{\mathbf{w}}_r + \sum_{s=1}^{j} b_s^{(i)}\hat{\mathbf{v}}_s. \tag{5.11}$$

By short recurrence of the Lanczos vectors, this implies

$$\mathbf{V}_{2k+\ell}\mathbf{T}_{2k+\ell}\mathbf{Q}\hat{\mathbf{q}}_i + \beta_{2k+\ell}\mathbf{v}_{2k+\ell+1}\mathbf{e}_{2k+\ell}^T\mathbf{Q}\hat{\mathbf{q}}_i = \mathbf{V}_{2k+\ell}\sum_{r=1}^{q} a_r^{(i)}\mathbf{Q}\hat{\mathbf{q}}_r + \sum_{s=1}^{j} b_s^{(i)}\hat{\mathbf{v}}_s,$$

where $\mathbf{T}_{2k+\ell} := \mathbf{V}_{2k+\ell}^T\mathbf{A}\mathbf{V}_{2k+\ell}$ The vectors $\hat{\mathbf{v}}_1, \dots, \hat{\mathbf{v}}_j$ being orthogonal to $\hat{\mathbf{w}}_1, \dots, \hat{\mathbf{w}}_q$, they are also orthogonal to the Lanczos vectors in $\mathbf{V}_{2k+\ell}$, so that we must have

$$\mathbf{V}_{2k+\ell}^T\mathbf{A}\mathbf{V}_{2k+\ell}\mathbf{Q}\hat{\mathbf{q}}_i = \sum_{r=1}^{q} a_r^{(i)}\mathbf{Q}\hat{\mathbf{q}}_r.$$

But, because $q > k$, there exists $\hat{\mathbf{q}}_i$ with $i \in [1, q]$ such that $\mathbf{Q}\hat{\mathbf{q}}_i$ is *not* an eigenvector of $\mathbf{V}_{2k+\ell}^T\mathbf{A}\mathbf{V}_{2k+\ell}$. Moreover, $\mathbf{A}$ being definite, $\ell > 0$ implies that $\mathcal{R}([\mathbf{Q}\hat{\mathbf{q}}_1, \dots, \mathbf{Q}\hat{\mathbf{q}}_q])$ does not contain $\mathcal{R}(\mathbf{T}_{2k+\ell})$. We say that $a_1^{(i)}, \dots, a_q^{(i)}$ can not exist for this $\hat{\mathbf{q}}_i$. ∎

A fortiori, letting $\hat{\mathbf{v}}_1 := \mathbf{v}_{2k+\ell+1}$—analogously to what was done in Section 5.1 when restarting with Raleigh-Ritz vectors computed in a single Krylov subspace—has for consequence that at least one vector $\hat{\mathbf{w}}_i$ among $\hat{\mathbf{w}}_1, \dots, \hat{\mathbf{w}}_q$ is such that no Krylov subspace $\mathcal{K}_d(\mathbf{A}, \hat{\mathbf{w}}_i)$ of dimension $d > 1$ is contained within $\mathcal{S}_{\text{LO-TR-Lan}}$. More importantly, it can also be shown that $\mathcal{S}_{\text{LO-TR-Lan}}$, in its whole, is not a Krylov subspace, although it does contain smaller Krylov subspaces. Therefore, using $\mathcal{S}_{\text{LO-TR-Lan}}(\{\hat{\mathbf{w}}_r\}_{r=1}^q, \hat{\mathbf{v}}_1, \ell)$ and $\mathcal{S}_{\text{LO-TR-Lan}}(\{\hat{\mathbf{w}}_r\}_{r=1}^q, \hat{\mathbf{v}}_1, \ell-1)$ to generate a search space and then compute new vectors $\hat{\mathbf{w}}_1, \dots, \hat{\mathbf{w}}_q$, has important consequences on the next restart. Indeed, the newly restarted subspace $\mathcal{S}_{\text{LO-TR-Lan}}$ will not contain any Krylov subspace of dimension $d > 1$ generated by either of the approximations $\hat{\mathbf{w}}_1, \dots, \hat{\mathbf{w}}_q$. Not only will this cause short recurrence relations to fail—thus increasing the cost of the Gram-Schmidt orthogonalization—but also, as our numerical results show, to severely hinder the convergence of this procedure.

Although it is not possible to set $\hat{\mathbf{v}}_1$ so that $\mathcal{S}_{\text{LO-TR-Lan}}$ is a Krylov subspace, it still can be reset for each restart so that $\mathcal{S}_{\text{LO-TR-Lan}}(\{\hat{\mathbf{w}}_r\}_{r=1}^q, \hat{\mathbf{v}}_1, j)$ contains a Krylov subspace $\mathcal{K}_{j+1}(\mathbf{A}, \hat{\mathbf{w}}_i)$ for some $\hat{\mathbf{w}}_i$. Assuming $\hat{\mathbf{w}}_1, \dots, \hat{\mathbf{w}}_q$ are ordered from least to most dominant, the following strategies are considered:

- *Least dominant eigen-residual*: Let $\hat{\mathbf{v}}_1 := \tilde{\mathbf{r}}_1/\|\tilde{\mathbf{r}}_1\|_2$, where $\tilde{\mathbf{r}}_i := \mathbf{A}\hat{\mathbf{w}}_i - \lambda_i\hat{\mathbf{w}}_i$.
- *Cyclic eigen-residual selection*: A new index $t \in [1, k]$ is picked for each restart, and $\hat{\mathbf{v}}_1 := \tilde{\mathbf{r}}_t/\|\tilde{\mathbf{r}}_t\|_2$.

While we expect the first strategy to yield a better convergence of the least dominant approximate eigen-pair, the latter should enable a better convergence of the whole approximated subspace. In either case, the cost of the Gram-Schmidt orthogonalization is the same, still affected by the fact that short recurrence relations do not hold after the first restart. Note that another alternative would be to let $\hat{\mathbf{v}}_1$ be a linear combination of eigen-residuals, this would however require to compute several eigen-residuals for each restart.

**5.2.2. Harmonic Rayleigh-Ritz.** Analogous procedures are implemented with harmonic Rayleigh-Ritz projections.

**5.3. Numerical results.** First, we present numerical results for TR-Lan in Fig. 3. Once again, only the 3 LD eigenvector approximations are considered, and the search space is restarted every time it reaches a dimension of 20. Clearly, restarting with $\hat{\mathbf{v}}_{k+1} := \mathbf{v}_{m+1}$ has serious consequences on the convergence of the harmonic procedure which, unlike Rayleigh-Ritz, stagnates both in terms of approximate eigenvectors and Rayleigh quotients. As explained before, this issue of harmonic approximations is solved when setting $\hat{\mathbf{v}}_{k+1} := \tilde{\mathbf{r}}$. However, once again, while leading to a faster convergence of eigen-residuals and Rayleigh quotients, the resulting procedure does not exhibit a faster convergence of the eigenvectors and the subspace they span, than the approximations obtained by Rayleigh-Ritz.

Second, corresponding results obtained for LO-TR-Lan are presented in Fig. 4. Now, just as explained before, letting $\hat{\mathbf{v}}_{q+1} := \mathbf{v}_{m+1}$ severely alters the behavior of both the harmonic and Rayleigh-Ritz procedures. Fortunately, letting $\hat{\mathbf{v}}_{q+1}$ be a selected eigen-residual makes up for these limitations. In particular, consistently picking the eigen-residual of the LD approximation, causes this pair to converge faster while the other approximations stagnate; and alternating between eigen-residuals from one restart to another, enables all the approximated eigenpairs to convergence simultaneously. Note that the effects of the locally optimal restart, in comparison to the behavior of TR-Lan, seem to appear once the Rayleigh quotients have converged. Then, once again, the harmonic procedure shows faster convergence of eigen-residuals and Rayleigh quotients. In both cases, this only translates to an increase in the convergence rate of the approximated eigenvectors once the Rayleigh quotients have converged.

**6. Recycling linear conjugate gradient solvers.** Consider a CG algorithm started with $\mathbf{x}_0$, and the iterates $\mathbf{x}_1, \mathbf{x}_2, \ldots$ generated as approximate solutions of $\mathbf{A}\mathbf{x} = \mathbf{b}$. The underlying sequence of residuals $\mathbf{r}_r := \mathbf{A}\mathbf{x}_r - \mathbf{b}$ is such that $\{\mathbf{v}_r\}_{r=1}^{j}$, in which $\mathbf{v}_r := (-1)^r\mathbf{r}_r/\|\mathbf{r}_r\|_2$, are the Lanczos vectors of $\mathcal{K}_j(\mathbf{A}, \mathbf{r}_0)$. By default, CG does not require to store these residuals, so that a recycling scheme remains to be defined in order to construct the basis of a search space within which eigenvectors of $\mathbf{A}$ are ought to be approximated. For the same reasons that thick-restarts are considered over un-restarted Lanczos methods, we want to sequentially re-define smaller search spaces rather than storing a large number of residual vectors. However, the difference with thick-restart strategies is that, here, the restarted search space is directly augmented with the vectors generated by the solver, whereas a thick-restart would generate this new sequence of vectors by orthogonalizing with respect to the approximations used to restart the search space.

**6.1. Thick recycling.** Let $\mathbf{w}_1, \ldots, \mathbf{w}_k$ be the $k$ LD eigenvector approximations in a search space $\mathcal{R}(\mathbf{V}_{k+\ell})$ where the columns of $\mathbf{V}_{k+\ell}$ are the Lanczos vectors obtained from the $(k+\ell)$ first

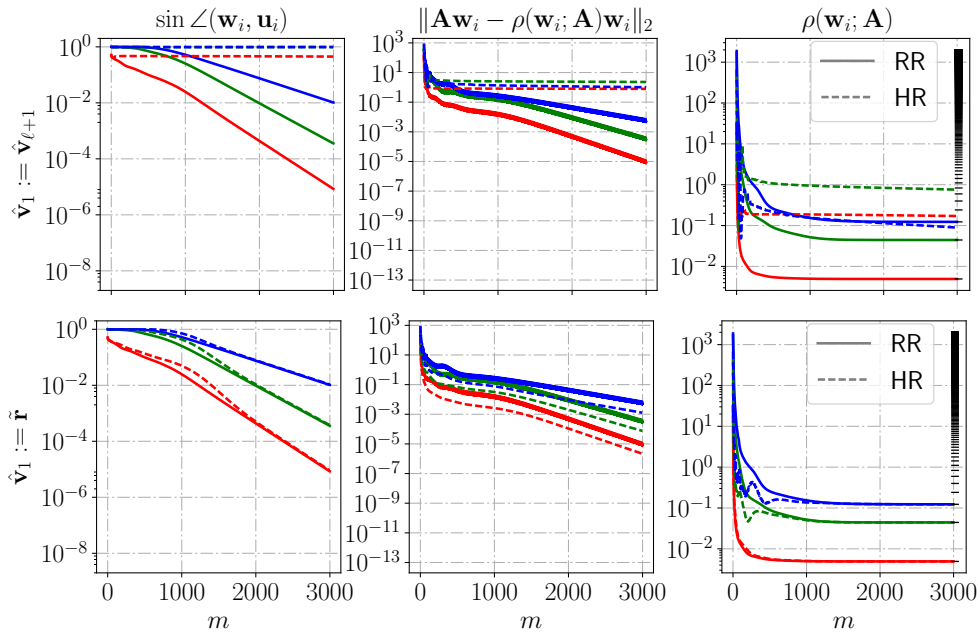


Figure 3: Results of thick-restart Lanczos procedures (TR-Lan) with RR/HR approximations and $\ell = 20$. The LD eigenvectors $\mathbf{u}_i$ of $\mathbf{A}$ are approximated by $\mathbf{w}_i$ for $i = 1$ (—), $i = 2$ (—), $i = 3$ (—). The black ticks on the third column represent the spectrum $\mathrm{Sp}(\mathbf{A})$.

residuals generated by the linear solver. Then, the most straightforward recycling strategy consists of using these $k$ approximations as a basis along with the Lanczos vectors which correspond to the very next residuals $\mathbf{r}_{k+\ell+1}, \mathbf{r}_{k+\ell+2}, \ldots$ of the linear system. Once $\ell$ new such vectors have been generated and stored, the so-formed basis is used to compute $k$ new eigenvector approximations. As this process is repeated over and over again, the current search space after $m + j$ solver iterations can be put in the form

$$\begin{aligned} \mathcal{S}_{\text{TR-CG}}(\{\mathbf{w}_r\}_{r=1}^{k}, j) &= \mathrm{Span}\{\mathbf{w}_1, \ldots, \mathbf{w}_k, \mathbf{v}_{m+1}, \ldots, \mathbf{v}_{m+j}\}, \\ &= \mathrm{Span}\{\mathbf{w}_1, \ldots, \mathbf{w}_k, \mathbf{v}_{m+1}, \mathbf{A}\mathbf{v}_{m+1} \ldots, \mathbf{A}^{j-1}\mathbf{v}_{m+1}\}, \end{aligned} \tag{6.1}$$

where $m$ is the number of linear solver iterations completed just before the last restart.

This procedure is implemented using both Rayleigh-Ritz and harmonic Rayleigh-Ritz projections. In both cases, it can be shown that the search spaces generated by this procedure are not Krylov. Also, no property analogous to the one given by Eq. (5.7) seem to hold, for either projection technique.

**6.2. Locally optimal recycling.** Similarly as for locally optimal thick-restart strategies, one could think of using eigenvector approximations from both the current and the previous search spaces. Therefore, we consider the following procedure. First, let $\mathbf{w}_1, \ldots, \mathbf{w}_k$ be the $k$ LD eigenvector approximations computed in the search space $\mathcal{R}(\mathbf{V}_{2k+\ell})$ where the columns of $\mathbf{V}_{2k+\ell}$ are the Lanczos vectors obtained from the $(2k + \ell)$ first residuals generated by the linear solver. Addition-

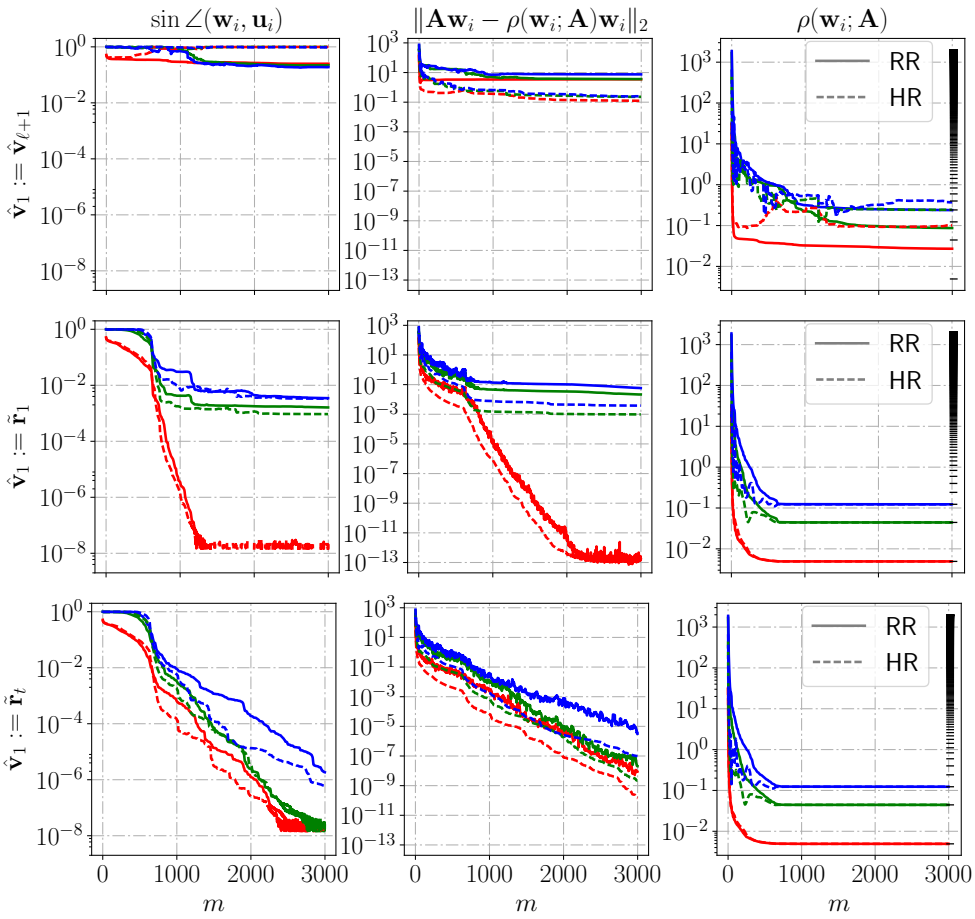


Figure 4: Results of locally optimal thick-restart Lanczos procedures (LO-TR-Lan) with RR/HR approximations and $\ell = 20$. The LD eigenvectors $\mathbf{u}_i$ of $\mathbf{A}$ are approximated by $\mathbf{w}_i$ for $i = 1$ (—), $i = 2$ (—), $i = 3$ (—). The black ticks on the third column represent the spectrum $\mathrm{Sp}(\mathbf{A})$.

ally, let $\overline{\mathbf{w}}_1, \ldots, \overline{\mathbf{w}}_k$ be the $k$ LD approximations in $\mathcal{R}(\mathbf{V}_{2k+\ell-1})$. Then, all these $2k$ approximations are used as a basis along with the Lanczos vectors which correspond to the very next residuals $\mathbf{r}_{2k+\ell+1}, \mathbf{r}_{2k+\ell+2}, \ldots$ of the linear system. Once $\ell$ new such vectors have been generated and stored, the so-formed basis is used to compute $k$ new eigenvector approximations from each of the current and previous search spaces. As this process is repeated over and over again, the current search space after $m + j$ solver iterations can be put in the form

$$\begin{aligned}\mathcal{S}_{\text{LO-TR-CG}}(\{(\mathbf{w}_r, \overline{\mathbf{w}}_r)\}_{r=1}^{k}, j) &= \mathrm{Span}\{\mathbf{w}_1, \overline{\mathbf{w}}_1, \ldots, \overline{\mathbf{w}}_k, \mathbf{w}_k, \mathbf{v}_{m+1}, \ldots, \mathbf{v}_{m+j}\}, \\ &= \mathrm{Span}\{\mathbf{w}_1, \overline{\mathbf{w}}_1, \ldots, \overline{\mathbf{w}}_k, \mathbf{w}_k, \mathbf{v}_{m+1}, \mathbf{A}\mathbf{v}_{m+1} \ldots, \mathbf{A}^{j-1}\mathbf{v}_{m+1}\},\end{aligned} \tag{6.2}$$

where $m$ is the number of linear solver iterations completed just before the last restart.

Just as it happens with locally optimal thick restarts, the pairs $(\mathbf{w}_1, \overline{\mathbf{w}}_1)$ through $(\mathbf{w}_k, \overline{\mathbf{w}}_k)$ tend to be co-linear. Consequently, the same solution is applied. For instance, at the first restart, instead of using $\mathbf{w}_1, \overline{\mathbf{w}}_1, \ldots, \mathbf{w}_k, \overline{\mathbf{w}}_k$ directly, we rather use all the $q$ eigenvector approximations $\hat{\mathbf{w}}_1, \ldots, \hat{\mathbf{w}}_q$ of the subspace $\mathcal{R}(\mathbf{V}_{2k+\ell}\mathbf{Q}) = \text{Span}\{\mathbf{w}_1, \overline{\mathbf{w}}_1, \ldots, \mathbf{w}_k, \overline{\mathbf{w}}_k\}$, where $\mathbf{Q}$ has full column rank $q$ with $k < q \leq 2k$. More particularly, the columns of $\mathbf{Q}$ form an orthonormal basis of

$$\mathcal{R}(\mathbf{Q}) = \text{Span}\left\{\mathbf{y}_1, \begin{bmatrix}\overline{\mathbf{y}}_1 \\ 0\end{bmatrix}, \ldots, \mathbf{y}_k, \begin{bmatrix}\overline{\mathbf{y}}_k \\ 0\end{bmatrix}\right\},$$

where $\mathbf{y}_1, \ldots, \mathbf{y}_k$ and $\overline{\mathbf{y}}_1, \ldots, \overline{\mathbf{y}}_k$ are the reduced eigenvectors such that $\mathbf{w}_i = \mathbf{V}_{2k+\ell}\mathbf{y}_i$ and $\overline{\mathbf{w}}_i = \mathbf{V}_{2k+\ell-1}\overline{\mathbf{y}}_i$, respectively.

**6.2.1. Rayleigh-Ritz.** The case in which all the pairs $(\mathbf{w}_1, \overline{\mathbf{w}}_1)$ through $(\mathbf{w}_k, \overline{\mathbf{w}}_k)$ are Rayleigh-Ritz approximations was originally proposed in [35], where it is referred to as eigCG, and used to solve sequences of linear systems with multiple right-hand sides. This method is known to exhibit convergence behaviors which are as good as unrestarted Lanczos procedures, although little seems to be known as to why [35, **?**]. We believe this is because, for each approximation $\mathbf{w}_i$, there exists some $\tilde{\mathbf{w}}_i$ such that $\|\mathbf{w}_i - \tilde{\mathbf{w}}_i\|_2 \ll 1$, and which satisfies

$$\mathcal{K}_{j+1}(\mathbf{A}, \tilde{\mathbf{w}}_i) \subset \mathcal{S}_{\text{LO-TR-CG}}(\{(\mathbf{w}_r, \overline{\mathbf{w}}_r)\}_{r=1}^k, j). \tag{6.3}$$

This property, in turn, is believed to strongly influence the convergence behavior of the procedure. To show that Eq. (6.3) holds, we consider $\mathcal{S}_{\text{LO-TR-CG}}(\{(\mathbf{w}_r, \overline{\mathbf{w}}_r)\}_{r=1}^k, \ell)$ where $\mathbf{w}_r := \mathbf{V}_m\mathbf{y}_r$ is the $r$-th LD Rayleigh-Ritz vector of $\mathbf{A}$ in $\mathcal{R}(\mathbf{V}_m)$, as opposed to $\overline{\mathbf{w}}_s := \mathbf{V}_{m-1}\overline{\mathbf{y}}_s$, which is the $s$-th one in $\mathcal{R}(\mathbf{V}_{m-1})$. Let $\mathbf{W} := [\mathbf{w}_1, \ldots, \mathbf{w}_k]$, $\overline{\mathbf{W}} := [\overline{\mathbf{w}}_1, \ldots, \overline{\mathbf{w}}_k]$ and $\hat{\mathbf{V}} := [\mathbf{W}, \overline{\mathbf{W}}, \mathbf{v}_{m+1}, \ldots, \mathbf{v}_{m+\ell}]$. Then, remarkably, even though $\mathcal{R}(\hat{\mathbf{V}})$ is not a Krylov subspace, it contains Rayleigh-Ritz pairs $(\vartheta, \hat{\mathbf{V}}\mathbf{z})$ of $\mathbf{A}$ with eigen-residual $\tilde{\mathbf{r}}(\vartheta, \mathbf{z}) := \mathbf{A}\hat{\mathbf{V}}\mathbf{z} - \vartheta\hat{\mathbf{V}}\mathbf{z}$ such that

$$\lim_{\|\mathbf{I}_k - \mathbf{W}^T\overline{\mathbf{W}}\|_2 \to 0} \tilde{\mathbf{r}}(\vartheta, \mathbf{z}) = \beta_{m+\ell+1}(\mathbf{e}_{2k+\ell}^T\mathbf{z})\hat{\mathbf{v}}_{m+\ell+1}. \tag{6.4}$$

*Proof.* Since the Rayleigh-Ritz pairs $(\lambda_r, \mathbf{w}_r)$ and $(\overline{\lambda}_s, \overline{\mathbf{w}}_s)$ are computed in Krylov search spaces, they satisfy the following equations:

$$\mathbf{A}\mathbf{w}_r = \lambda_r\mathbf{w}_r + \beta_m(\mathbf{e}_m^T\mathbf{y}_r)\mathbf{v}_{m+1} \tag{6.5}$$

$$\mathbf{A}\overline{\mathbf{w}}_s = \overline{\lambda}_s\overline{\mathbf{w}}_s + \beta_{m-1}(\mathbf{e}_{m-1}^T\overline{\mathbf{y}}_s)\mathbf{v}_m \tag{6.6}$$

where $\mathbf{e}_\bullet$ is the last column of the $\bullet$-dimensional identity matrix. Meanwhile, the vectors $\mathbf{v}_{m+1}, \ldots, \mathbf{v}_{m+\ell}$ satisfy the following short recurrence relation:

$$\mathbf{A}\mathbf{v}_{m+j} = \alpha_{m+j}\mathbf{v}_{m+j} + \beta_{m+j+1}\mathbf{v}_{m+j+1} + \beta_{m+j}\mathbf{v}_{m+j-1}. \tag{6.7}$$

Eqs. (6.5) through (6.7) can then be recast into the following matrix form:

$$\mathbf{A}\hat{\mathbf{V}} = \hat{\mathbf{V}}\hat{\mathbf{M}} + \sum_{s=1}^{k} \beta_{m-1}(\mathbf{e}_{m-1}^T\overline{\mathbf{y}}_s)\mathbf{v}_m\hat{\mathbf{e}}_{k+s}^T + \beta_{m+1}\mathbf{v}_m\hat{\mathbf{e}}_{2k+1}^T + \beta_{m+\ell+1}\mathbf{v}_{m+\ell+1}\hat{\mathbf{e}}_{2k+\ell}^T, \tag{6.8}$$

where $\hat{\mathbf{e}}_\bullet$ is the $\bullet$-th column of the $(2k+\ell)$-dimensional identity matrix, and

$$\begin{aligned}\hat{\mathbf{M}} := &\sum_{r=1}^{k} \lambda_r \hat{\mathbf{e}}_r \hat{\mathbf{e}}_r^T + \sum_{s=1}^{k} \overline{\lambda}_s \hat{\mathbf{e}}_{k+s} \hat{\mathbf{e}}_{k+s}^T + \sum_{r=1}^{k} \beta_m (\mathbf{e}_m^T \mathbf{y}_r) \hat{\mathbf{e}}_{2k+1} \hat{\mathbf{e}}_r^T \\ &+ \sum_{r=1}^{k} \alpha_{m+j} \hat{\mathbf{e}}_{2k+j} \hat{\mathbf{e}}_{2k+j}^T + \sum_{r=1}^{k} \beta_{m+j} \left( \hat{\mathbf{e}}_{2k+j} \hat{\mathbf{e}}_{2k+j-1}^T + \hat{\mathbf{e}}_{2k+j-1} \hat{\mathbf{e}}_{2k+j}^T \right).\end{aligned} \tag{6.9}$$

Note that $\mathbf{A}\hat{\mathbf{V}}$ can not be put in the form of a rank-1 update, so that $\mathcal{R}(\hat{\mathbf{V}})$ is not a Krylov subspace [37]. Consequently, the usual expressions for the eigen-residuals of Rayleigh-Ritz pairs in Krylov search spaces do not apply to approximations in $\mathcal{R}(\hat{\mathbf{V}})$. To find expressions for those eigen-residuals, we first search an expression for $\hat{\mathbf{T}} := \hat{\mathbf{V}}^T \mathbf{A} \hat{\mathbf{V}}$. Using Eq. (6.8)–(6.9), we get:

$$\begin{aligned}\hat{\mathbf{T}} = \hat{\mathbf{V}}^T (\mathbf{A}\hat{\mathbf{V}}) = \hat{\mathbf{M}} &+ \sum_{r=1}^{k} \sum_{s=1}^{k} \lambda_r (\mathbf{w}_r^T \overline{\mathbf{w}}_s) \hat{\mathbf{e}}_{k+s} \hat{\mathbf{e}}_r^T + \sum_{r=1}^{k} \sum_{s=1}^{k} \overline{\lambda}_s (\mathbf{w}_r^T \overline{\mathbf{w}}_s) \hat{\mathbf{e}}_r \hat{\mathbf{e}}_{k+s}^T \\ &+ \sum_{r=1}^{k} \sum_{s=1}^{k} \beta_{m-1} (\mathbf{e}_m^T \mathbf{y}_r)(\mathbf{e}_{m-1}^T \overline{\mathbf{y}}_s) \hat{\mathbf{e}}_r \hat{\mathbf{e}}_{k+s}^T + \sum_{r=1}^{k} \beta_{m+1} (\mathbf{e}_m^T \mathbf{y}_r) \hat{\mathbf{e}}_r \hat{\mathbf{e}}_{2k+1}^T.\end{aligned} \tag{6.10}$$

Note that the approximations $\mathbf{w}_1, \ldots, \mathbf{w}_k$ and $\overline{\mathbf{w}}_1, \ldots, \overline{\mathbf{w}}_k$ are linearly independent, but not orthogonal, i.e. $\mathbf{w}_r^T \overline{\mathbf{w}}_s \neq 0$. Therefore, the Rayleigh-Ritz pair $(\vartheta, \hat{\mathbf{V}}\mathbf{z})$ of $\mathbf{A}$ in $\mathcal{R}(\hat{\mathbf{V}})$ is such that $\hat{\mathbf{T}}\mathbf{z} = \vartheta \hat{\mathbf{V}}^T \hat{\mathbf{V}} \mathbf{z}$. Then, using Eqs. (6.8)–(6.10), we obtain:

$$\begin{aligned}\mathbf{A}\hat{\mathbf{V}}\mathbf{z} = \vartheta \hat{\mathbf{V}}\mathbf{z} &+ \beta_{m+\ell+1} (\hat{\mathbf{e}}_{2k+\ell}^T \mathbf{z}) \mathbf{v}_{m+\ell+1} \\ &+ \sum_{s=1}^{k} \beta_{m-1} (\mathbf{e}_{m-1}^T \overline{\mathbf{y}}_s)(\hat{\mathbf{e}}_{k+s}^T \mathbf{z}) \mathbf{v}_m + \beta_{m+1} (\mathbf{e}_{2k+1}^T \mathbf{z}) \mathbf{v}_m \\ &- \sum_{r=1}^{k} \sum_{s=1}^{k} \lambda_r (\mathbf{w}_r^T \overline{\mathbf{w}}_s)(\hat{\mathbf{e}}_r^T \mathbf{z}) \overline{\mathbf{w}}_s - \sum_{r=1}^{k} \sum_{s=1}^{k} \overline{\lambda}_s (\mathbf{w}_r^T \overline{\mathbf{w}}_s)(\hat{\mathbf{e}}_{k+s}^T \mathbf{z}) \mathbf{w}_r \\ &- \sum_{r=1}^{k} \sum_{s=1}^{k} \beta_{m-1} (\mathbf{e}_m^T \mathbf{y}_r)(\mathbf{e}_{m-1}^T \overline{\mathbf{y}}_s)(\hat{\mathbf{e}}_{k+s}^T \mathbf{z}) \mathbf{w}_r - \sum_{r=1}^{k} \beta_{m+1} (\mathbf{e}_m^T \mathbf{y}_r)(\hat{\mathbf{e}}_{2k+1}^T \mathbf{z}) \mathbf{w}_r.\end{aligned} \tag{6.11}$$

By definition of Rayleigh-Ritz pairs, the eigen-residual $\tilde{\mathbf{r}}(\vartheta, \mathbf{z}) := \mathbf{A}\hat{\mathbf{V}}\mathbf{z} - \vartheta \hat{\mathbf{V}}\mathbf{z}$ is orthogonal to $\mathcal{R}(\hat{\mathbf{V}})$. Therefore, applying $\hat{\mathbf{V}}^T \tilde{\mathbf{r}}(\vartheta, \mathbf{z}) = 0$ to Eq. (6.11), we get:

$$\overline{\lambda}_s (\hat{\mathbf{e}}_{k+s}^T \mathbf{z}) = -\sum_{r=1}^{k} \lambda_r (\mathbf{w}_r^T \overline{\mathbf{w}}_s)(\hat{\mathbf{e}}_r^T \mathbf{z}) \quad \text{for} \quad 1 \leq s \leq k, \tag{6.12}$$

as well as

$$\begin{aligned}\lambda_r (\hat{\mathbf{e}}_r^T \mathbf{z}) + \sum_{s=1}^{k} \overline{\lambda}_s (\mathbf{w}_r^T \overline{\mathbf{w}}_s)(\hat{\mathbf{e}}_{k+s}^T \mathbf{z}) = &- \sum_{s=1}^{k} \beta_{m-1} (\mathbf{e}_m^T \mathbf{y}_r)(\mathbf{e}_{m-1}^T \overline{\mathbf{y}}_s)(\hat{\mathbf{e}}_{k+s}^T \mathbf{z}) \\ &- \beta_{m+1} (\mathbf{e}_m^T \mathbf{y}_r)(\hat{\mathbf{e}}_{2k+1}^T \mathbf{z})\end{aligned} \tag{6.13}$$

for $1 \leq r \leq k$. Using Eqs. (6.12)–(6.13) into Eq. (6.11), we obtain the following expression for the eigen-residual:

$$\begin{aligned}\tilde{\mathbf{r}}(\vartheta, \mathbf{z}) = \beta_{m+\ell+1}(\hat{\mathbf{e}}_{2k+\ell}^T \mathbf{z})\mathbf{v}_{m+\ell+1} + \sum_{r=1}^{k} \lambda_r(\hat{\mathbf{e}}_r^T \mathbf{z}) \left( \mathbf{w}_r - \sum_{s=1}^{k} (\mathbf{w}_r^T \overline{\mathbf{w}}_s)\overline{\mathbf{w}}_s \right) \\ + \left( \beta_{m+1}(\hat{\mathbf{e}}_{2k+1}^T \mathbf{z}) + \sum_{s=1}^{k} \beta_{m-1}(\mathbf{e}_{m-1}^T \overline{\mathbf{y}}_s)(\hat{\mathbf{e}}_{k+s}^T \mathbf{z}) \right) \mathbf{v}_m.\end{aligned} \tag{6.14}$$

However, when $\mathbf{w}_r^T \overline{\mathbf{w}}_s \approx \delta_{rs}$, i.e. as $\|\mathbf{I}_k - \mathbf{W}^T \overline{\mathbf{W}}\|_2$ goes to zero, we have

$$\mathbf{w}_r - \sum_{s=1}^{k} (\mathbf{w}_r^T \overline{\mathbf{w}}_s)\overline{\mathbf{w}}_s \approx \mathbf{w}_r - (\mathbf{w}_r^T \overline{\mathbf{w}}_r)\overline{\mathbf{w}}_r \approx \mathbf{w}_r - \overline{\mathbf{w}}_r \approx 0.$$

Similarly, using Eq. (6.12), we can see that the left-hand side of Eq. (6.13) nearly cancels when $\|\mathbf{I}_k - \mathbf{W}^T \overline{\mathbf{W}}\|_2 \approx 0$, in which case the term multiplying $\mathbf{v}_m$ in Eq. (6.14) also vanishes. ∎

Note that, as mentioned earlier, we do observe $\mathbf{w}_i^T \overline{\mathbf{w}}_j \approx \delta_{ij}$ directly after the first restart. Letting $\mathbf{w}_i$ be a Rayleigh-Ritz approximation in a search space between the first and second restarts, there exists $\tilde{\mathbf{w}}_i$ such that $\|\mathbf{w}_i - \tilde{\mathbf{w}}_i\|_2 \ll 1$, and for which we have $\tilde{\mathbf{r}}(\vartheta, \tilde{\mathbf{w}}_i) \propto \hat{\mathbf{v}}_{m+\ell+1}$. This, in turn, implies Eq. (6.3). Note that, although the search space is not Krylov, this property remains valid after further restarts.

**6.2.2. Harmonic Rayleigh-Ritz.** The same procedure can be applied with harmonic Rayleigh-Ritz approximations. However, it does not seem to admit the same property.

**6.3. Numerical results.** TR-CG and LO-TR-CG are tested on a sequence of linear systems $\mathbf{A}\mathbf{x}^{(t)} = \mathbf{b}^{(t)}$ with random $\mathbf{b}^{(t)}$. Every solver run is initialized with an iterate $\mathbf{x}_0 = 0$. The forward error of the eigenvector, the eigen-residual and the Ritz quotients of the 3 LD approximations obtained while recycling the matrix-vector products of four iterative system resolutions, are presented in Fig. 5. Note that both harmonic and standard Rayleigh-Ritz projections are used to generate the results, for which the search space is restarted every time its dimension reaches 20. Clearly, the LO-TR-CG procedure based on a Rayleigh-Ritz projection is the only one which seems to converge within the first couple of systems. Actually, it does so so well, that it closely matches the behavior of the (orthogonalized) unrestarted Lanczos procedure, see Fig. 2. Note that this approach, when initialized properly, corresponds to eigCG, which was introduced by [35].

**6.3.1. Deflation of linear systems with multiple right hand sides.** The same sequence of linear systems is considered, and extended in order to illustrate a longer term effect of recycling strategies, on deflation. The same procedures are considered. Only this time, a Deflated-CG solver is used in place of CG. At the end of each solver run, the 10 LD eigenvector approximations are used as a basis of the deflation subspace used for the resolution of the next system in the sequence. We plot the iterated residuals of 50 linear systems in Fig. 6, along with corresponding results obtained by Init-CG, which is, simply by not forcing the iterated residual to remain orthogonal to the deflation subspace. For each procedure tested, the search space is restarted every time its dimension reaches 40. Note that, since LO-TR-CG starts every search space with more vectors (up to 20 in this case), it needs to perform more restarts than TR-CG. Not only are these restarts more frequent, they also require the resolution of two reduced eigenvalue problems, instead of one,

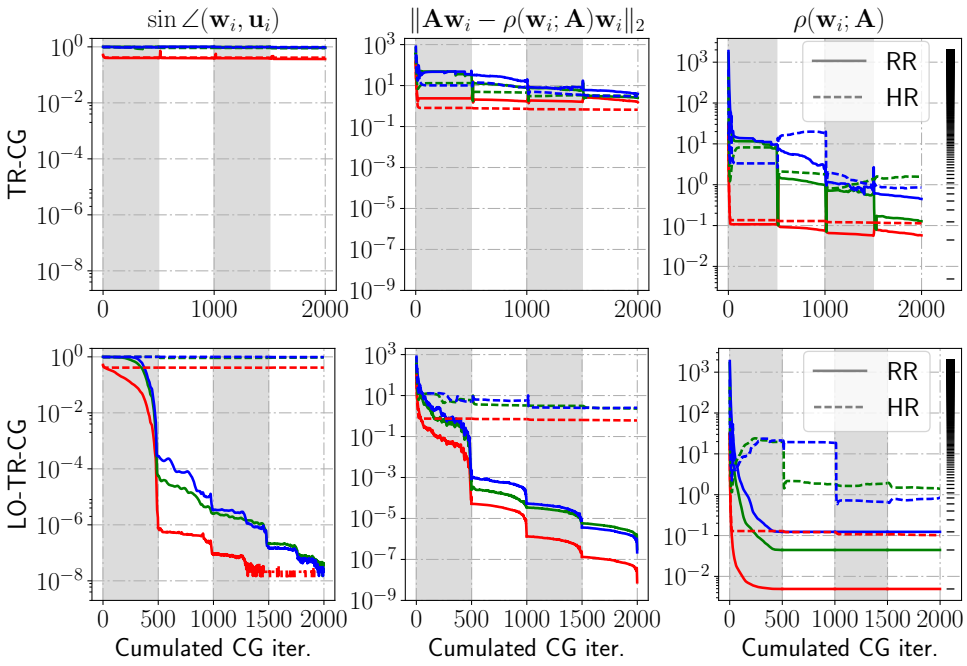


Figure 5: Results of TR-CG and LO-TR-CG procedures with RR and HR approximations. The LD eigenvectors $\mathbf{u}_i$ of $\mathbf{A}$ are approximated by $\mathbf{w}_i$ for $i = 1$ (—), $i = 2$ (—), $i = 3$ (—). The black ticks on the third column represent the spectrum $\mathrm{Sp}(\mathbf{A})$.

and a little more work than the restarts of TR-CG. The restarts based on harmonic Rayleigh-Ritz projections are also more demanding.

Clearly, the first row of Fig. 6 shows that the Rayleigh-Ritz version of LO-TR-Deflated-CG performs better than all the other procedures, as its effect on convergence is already completely achieved by the resolution of the second system. Although the results are not presented here, we tested the harmonic version of LO-TR-Deflated-CG, and it did not perform much better than the corresponding LO-TR-Deflated-CG procedure. Now, while TR-Deflated-CG procedures seem to require more systems to completely show their effect on convergence, they do yield similar accelerations. Note however that only LO-TR-Init-CG seems to reach approximations which are accurate enough so that Init-CG behaves similarly to Deflated-CG.

**6.3.2. Deflation of linear systems with multiple operators.** The same test is performed again, only this time, the right-hand side stays unchanged while $\mathbf{A}^{(t)}$ is randomly sampled after a random walk. The results presented in Fig. 7 are similar to those obtained for the sequence with multiple right-hand sides. Indeed, the Rayleigh-Ritz version of LO-TR-Deflated-CG shows quickly an effect on convergence, after a couple of systems have been solved, while it takes more systems for TR-Deflated-CG to reach the same acceleration. However, none of these strategies seem to enable a robust use of Init-CG for stopping criteria under $10^{-2} \times \|\mathbf{b}\|_2$.

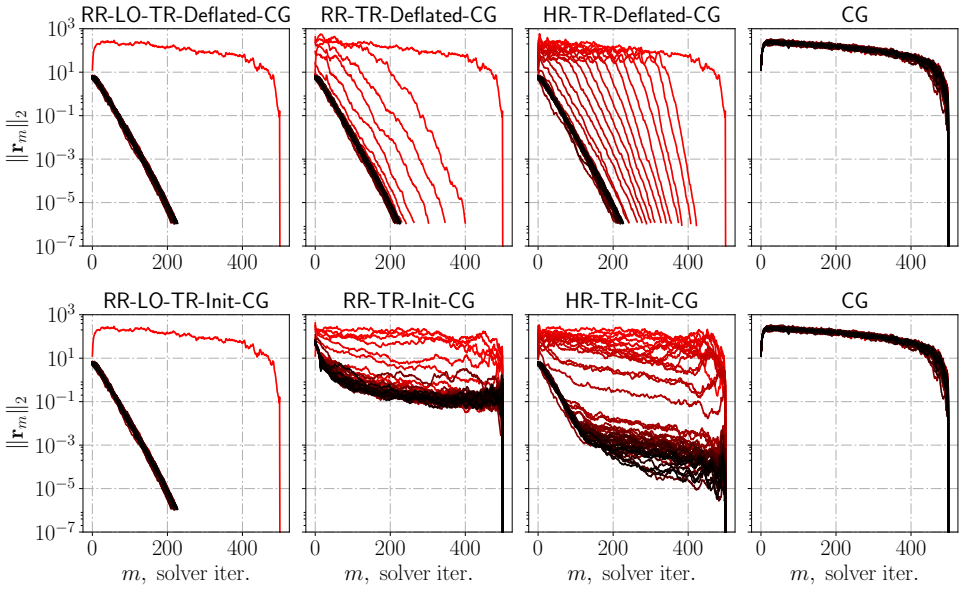


Figure 6: Norm of iterated residuals of recycled Deflated- and Init-CG procedures for a sequence of linear systems with multiple right-hand sides. The search spaces are restarted every time their dimension reaches 40. The reddest and blackest curves are for the first and last systems solved, respectively.

**7. Conclusion.** Harmonic Rayleigh-Ritz and Raleigh-Ritz projection techniques were applied to (i) locally optimal conjugate gradient (CG) methods, (ii) thick-restart Lanczos methods, and (iii) recycled linear CG solvers. Attempts were made to adapt approaches based on principles of local optimality and enable them to be used with harmonic projection techniques. While it seems that LOBCG can be adapted to allow for harmonic projections, there remains to see if this is the case when (i) introducing a shift so as to approximate eigen-pairs in the center of the spectrum, (ii) using a preconditioner. Meanwhile, locally optimal thick-restarts were successfully applied to accelerate restarted Lanczos procedures based on both harmonic and Rayleigh-Ritz approximations. However, the search spaces generated are not Krylov, and the orthogonalization, which is an intrinsic part of restarting, does not simplify. Perhaps the most surprising is that, locally optimal recycled CG, when based on Rayleigh-Ritz projections as in [35], generates search spaces which nearly contain all the Krylov subspaces of all the eigenvector approximations, up to some dimension, this while simply recycling matrix-vector products of the solver, with no further orthogonalization. This property, also used to explain the efficiency of thick-restart procedures, justifies why eigCG works so well as a recycling strategy in comparison to the other procedures attempted. Nevertheless, while thick-recycling procedures need to be applied to more systems before their effect on convergence fully develops, they do seem to reach similar behaviors as LO-TR-CG. In general, it seems that procedures which rely on harmonic Rayleigh-Ritz projection techniques yield faster convergence of eigen-residuals than corresponding methods based on Rayleigh-Ritz projections. However, the forward error of the approximated eigenvectors only start converging faster once the Rayleigh

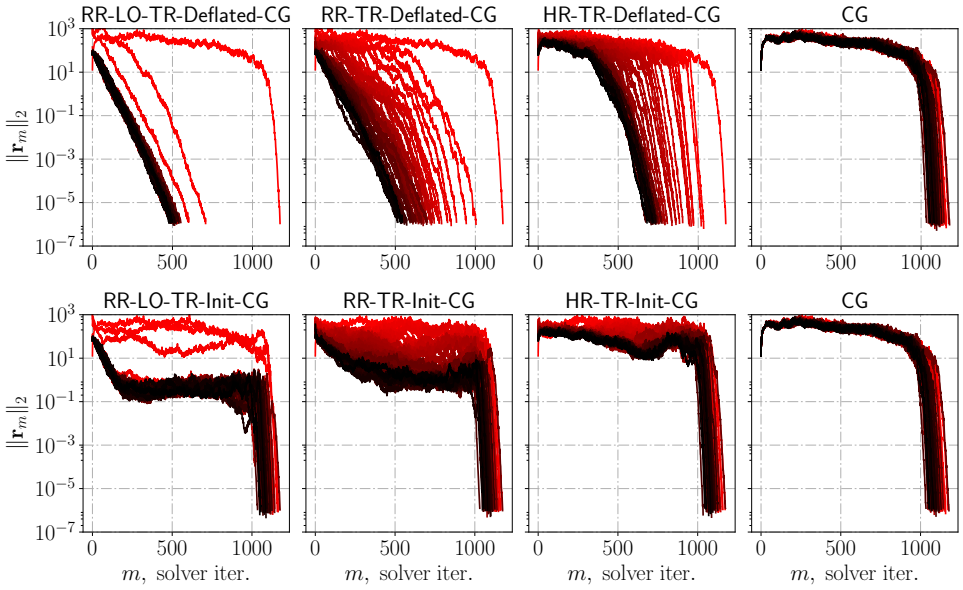


Figure 7: Norm of iterated residuals of recycled Deflated- and Init-CG procedures for a sequence of linear systems with multiple operators (same right-hand side). The search spaces are restarted every time their dimension reaches 40. The reddest and blackest curves are for the first and last systems solved, respectively.

quotients have converged.

## REFERENCES


[1] P.-A. Absil, R. Mahony, R. Sepulchre, and P. Van Dooren, A Grassmann–Rayleigh quotient iteration for computing invariant subspaces, SIAM review, 44 (2002), pp. 57–73.

[2] O. Coulaud, L. Giraud, P. Ramet, and X. Vasseur, Deflation and augmentation techniques in Krylov subspace methods for the solution of linear systems, arXiv preprint arXiv:1303.5692, (2013).

[3] E. G. D'yakonov, Iteration methods in eigenvalue problems, Mathematical Notes, 34 (1983), pp. 945–953.

[4] A. Edelman, T. A. Arias, and S. T. Smith, The geometry of algorithms with orthogonality constraints, SIAM journal on Matrix Analysis and Applications, 20 (1998), pp. 303–353.

[5] Y. Feng and D. Owen, Conjugate gradient methods for solving the smallest eigenpair of large symmetric eigenvalue problems, International Journal for Numerical Methods in Engineering, 39 (1996), pp. 2209–2229.

[6] L. Giraud, D. Ruiz, and A. Touhami, A comparative study of iterative solvers exploiting spectral information for SPD systems, SIAM Journal on Scientific Computing, 27 (2006), pp. 1760–1786.

[7] M. H. Gutknecht, Spectral deflation in Krylov solvers: A theory of coordinate space based methods, Electron. Trans. Numer. Anal, 39 (2012), pp. 156–185.

[8] K. Kahl and H. Rittich, The deflated conjugate gradient method: Convergence, perturbation and accuracy, Linear Algebra and its Applications, 515 (2017), pp. 111–129.

[9] A. Knyazev, A preconditioned conjugate gradient method for eigenvalue problems and its implementation in a subspace, in Numerical Treatment of Eigenvalue Problems Vol. 5/Numerische Behandlung von Eigenwertaufgaben Band 5, Springer, 1991, pp. 143–154.

[10] A. V. Knyazev, Toward the optimal preconditioned eigensolver: Locally optimal block preconditioned

conjugate gradient method, SIAM journal on scientific computing, 23 (2001), pp. 517–541.
[11] J. W. S. Lord Rayleigh, On the calculation of the frequency of vibration of a system in its gravest mode, with an example from hydrodynamics, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 47 (1899), pp. 566–572.
[12] R. Morgan, On restarting the Arnoldi method for large nonsymmetric eigenvalue problems, Mathematics of Computation, 65 (1996), pp. 1213–1230.
[13] R. B. Morgan, Computing interior eigenvalues of large matrices, Linear Algebra and its Applications, 154-156 (1991), pp. 289 – 309.
[14] R. B. Morgan, Implicitly restarted GMRES and Arnoldi methods for nonsymmetric systems of equations, SIAM Journal on Matrix Analysis and Applications, 21 (2000), pp. 1112–1135.
[15] R. B. Morgan and M. Zeng, Harmonic projection methods for large non-symmetric eigenvalue problems, Numerical linear algebra with applications, 5 (1998), pp. 33–55.
[16] R. B. Morgan and M. Zeng, A harmonic restarted Arnoldi algorithm for calculating eigenvalues and determining multiplicity, Linear algebra and its applications, 415 (2006), pp. 96–113.
[17] C. C. Paige, The computation of eigenvalues and eigenvectors of very large sparse matrices, PhD thesis, University of London, 1971.
[18] C. C. Paige, Computational variants of the lanczos method for the eigenproblem, IMA Journal of Applied Mathematics, 10 (1972), pp. 373–381.
[19] C. C. Paige, B. N. Parlett, and H. A. Van der Vorst, Approximate solutions and eigenvalue bounds from Krylov subspaces, Numerical linear algebra with applications, 2 (1995), pp. 115–133.
[20] M. L. Parks, E. De Sturler, G. Mackey, D. D. Johnson, and S. Maiti, Recycling Krylov subspaces for sequences of linear systems, SIAM Journal on Scientific Computing, 28 (2006), pp. 1651–1674.
[21] B. N. Parlett, A new look at the Lanczos algorithm for solving symmetric systems of linear equations, Linear algebra and its applications, 29 (1980), pp. 323–346.
[22] B. N. Parlett, The symmetric eigenvalue problem, vol. 20, SIAM, 1998.
[23] B. N. Parlett and D. S. Scott, The Lanczos algorithm with selective orthogonalization, Mathematics of computation, 33 (1979), pp. 217–238.
[24] W. Ritz, Über eine neue methode zur lösung gewisser variationsprobleme der mathematischen physik, Journal für die reine und angewandte Mathematik, 135 (1909), pp. 1–61.
[25] Y. Saad, On the Lanczos method for solving symmetric linear systems with several right-hand sides, Mathematics of computation, 48 (1987), pp. 651–662.
[26] Y. Saad, Numerical methods for large eigenvalue problems: revised edition, vol. 66, SIAM, 2011.
[27] Y. Saad, M. Yeung, J. Erhel, and F. Guyomarc'h, A deflated version of the conjugate gradient algorithm, SIAM J. Sci. Comput., 21 (1999), pp. 1909–1926.
[28] A. Sameh and Z. Tong, The trace minimization method for the symmetric generalized eigenvalue problem, Journal of Computational and Applied Mathematics, 123 (2000), pp. 155–175.
[29] D. Scott, The advantages of inverted operators in Rayleigh-Ritz approximations, SIAM Journal on Scientific and Statistical Computing, 3 (1982), pp. 68–75.
[30] H. D. Simon, The Lanczos algorithm with partial reorthogonalization, Mathematics of computation, 42 (1984), pp. 115–142.
[31] G. L. Sleijpen and H. A. Van der Vorst, A Jacobi–Davidson iteration method for linear eigenvalue problems, SIAM review, 42 (2000), pp. 267–293.
[32] D. C. Sorensen, Implicit application of polynomial filters in a k-step Arnoldi method, SIAM journal on matrix analysis and applications, 13 (1992), pp. 357–385.
[33] A. Stathopoulos, Nearly optimal preconditioned methods for Hermitian eigenproblems under limited memory. part I: Seeking one eigenvalue, SIAM Journal on Scientific Computing, 29 (2007), pp. 481–514.
[34] A. Stathopoulos and J. R. McCombs, Nearly optimal preconditioned methods for Hermitian eigenproblems under limited memory. part II: Seeking many eigenvalues, SIAM Journal on Scientific Computing, 29 (2007), pp. 2162–2188.
[35] A. Stathopoulos and K. Orginos, Computing and deflating eigenvalues while solving multiple right-hand side linear systems with an application to quantum chromodynamics, SIAM Journal on Scientific Computing, 32 (2010), pp. 439–462.
[36] A. Stathopoulos and Y. Saad, Restarting techniques for (Jacobi-) Davidson symmetric eigenvalue methods, Electr. Trans. Numer. Algebra, 7 (1998), p. 163181.
[37] G. W. Stewart, Matrix algorithms volume 2: eigensystems, vol. 2, SIAM, 2001.
[38] G. W. Stewart, A Krylov–Schur algorithm for large eigenproblems, SIAM Journal on Matrix Analysis and Applications, 23 (2002), pp. 601–614.
[39] H. A. Van der Vorst, An iterative solution method for solving f(A)x=b, using Krylov subspace information

obtained for the symmetric positive definite matrix A, Journal of Computational and Applied Mathematics, 18 (1987), pp. 249–263.

[40] K. Wu and H. Simon, Thick-restart Lanczos method for large symmetric eigenvalue problems, SIAM Journal on Matrix Analysis and Applications, 22 (2000), pp. 602–616.